

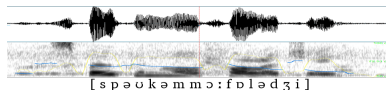
# Analogy in the Plural System of Maltese

Ruben van de Vijver (& Jessica Nieder)

Heinrich-Heine-Universität Düsseldorf

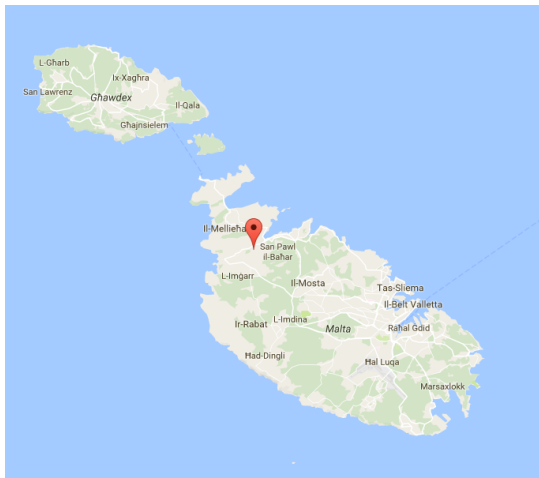
May 11, 2017

Linguistics Colloquium Tel Aviv University



# Maltese

Spoken here (courtesy google maps):



This is work in progress: There will be (many) rough edges and the paint isn't dry yet!

# Maltese

- It is a semitic language, with characteristics of Maghrebi Arabic and traces of Levantine Arabic.
- National language of Malta.
- Spoken by about 400.000 people in Malta (Malta, Gozo and 1 family in Comino).
- Another 100.000 people speak it around the world (Australia, the US, Canada, Belgium, Luxembourg, Italy and the UK.)

# Maltese

- It has been influenced by Italian (Sicilian) and English.
- The lexicon consists of 32% Arabic, 52% Italian and 6% English items. (And a rest of obscure origin (Brincat, 1996).)
  - ħabib 'friend'
  - furketta 'fork'
  - xawer 'shower'
- The Arabic words are most frequently used.

# Sound and broken plurals

**sound** add a suffix: sptar – sptar-ijiet 'hospital'

**broken** change the prosody: ktieb – kotba 'book'

# Sound plurals

Singular	Plural	Suffix	Gloss
arloġġ	arloġġi	-i	watch, clock
omm	ommiji:t	-iji:t	mother
ħaddi:m	ħaddi:ma	-a	worker
bnidem	bnedmi:n	-i:n	lazy
film	films	-s	movie
saltna	saltni:t	-a:t, -i:t	kingdom

Mayer, Spagnol & Schönhuber (2013)

# Broken plurals

Type	Singular	Plural	Gloss
A	bandi:ra	bna:dar	flag
B	balla	balal	bundle
C	borġ	bra:ġ	heap
D	xmara	xmajjar	river
E	xatba	xta:bi	gate
F	baħar	ibħra	sea
G	ġdid	ġdodda	new
H	għarbi [arbi]	għarab [arap]	Arab
I	wiċċ	uċuħ	face
J	għaref [aref] [aref]	għori:f [ori:f]	wise man
K	għama [ama]	għomja [omja]	blind person

Schembri (2012)

# Several sound plurals for one singular

Singular	Plural	Gloss
werqa	werq-at	leaf
werqa	werq-i:t	leaf

# Both sound and broken forms

for one singular

Singular	Broken plural	Sound plural	gloss
bandi:ra	bna:dar	bandi:ri	flag
tapit	twapet	tapiti	carpet
ħaxix	ħxejjex	ħaxixi:t	vegetables

# Sound? Broken? Bround!

Some forms seem to have both a suffix and a changed prosody:

Singular	Plural	Gloss
bni:dem	bnedm-i:n	lazy
giddi:b	giddib-in	liar (bround)

Borġ & Azzopardi-Alexander (1997)

# Maltese

## Great deal of variation

There is a great deal of variation. It is difficult to pinpoint the rules for sound plurals (Borġ & Azzopardi-Alexander, 1997), and broken plurals drive scholars of Maltese to despair: *"Dwar il-plural miksura m'hemmx regoli"*. (*There are no rules governing the broken plural*. (L-Għaqda Tal-Kittieba Maltin. Cited in: Schembri, 2012)

# Maltese

- If it is indeed the case that there are no rules governing the broken plural, this means that there is no – linguistic or statistical – structure in the data that allows native speakers to generalize.
- Broken plurals should not be productive.

# Maltese

## No unmarked shapes

The first syllable of many broken plurals have this shape:

CCVV

(for example: bnaːdar, braːġ.) This is not, as far as we know, an unmarked prosodic shape.

# Maltese

## Extant accounts

- prosodic morphology
  - Plural forms are not prosodically optimizing, nor are they prosodically unmarked.
- CV-skeleton mapping
  - What skeletons are chosen when?

# Maltese

## Extant accounts

- prosodic morphology
  - Plural forms are not prosodically optimizing, nor are they prosodically unmarked.
- CV-skeleton mapping
  - What skeletons are chosen when?
- The general idea behind these theories: the phonotactics of the singular determines the shape of the plural. This is a good idea.

# Maltese

## Hypothesis

- The phonotactics of the singular determines the shape of the plural.
- More frequent items are more likely to be generalized than infrequent items.

# Maltese

## Corpus and experiment

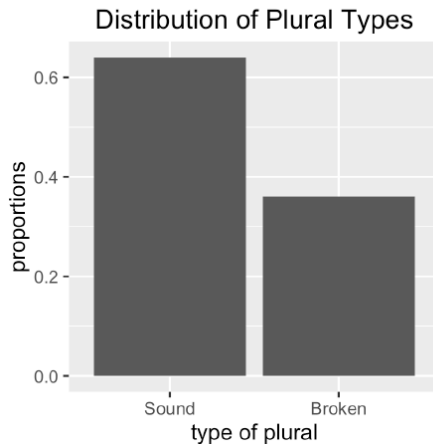
To test these hypotheses we created a corpus and we did a production experiment.

# Maltese

## Corpus

- We created a corpus of 2225 Maltese nouns
  - Taken from the online corpus MLRS Corpus Malti.
  - The corpus was checked by means of the online dictionary ġabra.

# Plurals in our corpus



# Maltese experiment

- We created nonce forms based on the forms found in our 2225 word corpus
  - We changed C or V or both systematically:
    - sema 'sky' → fera, soma, fora.
- We divided the words up in frequent ( $> 50$  per million) and infrequent ( $< 50$  per million).
- We chose 90 nonces (30 C-changed words, 30 V-changed words and 30 CV-words.)
- and 22 existing nouns:
  - 5 frequent sound plural words, 5 infrequent ones
  - 5 frequent broken plurals, 5 infrequent ones
  - 2 training items (1 sound, 1 broken.)

# Experiment

- Production test with visual presentation
- Software SpeechRecorder
- 38 native speakers of Maltese tested in Malta.
- First one item: *Dik l-stampa ta' X. This is a picture of X*
- Then 3 items: *ħafna X? Many X?*



# Results

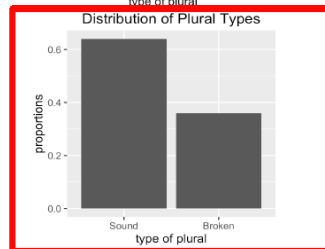
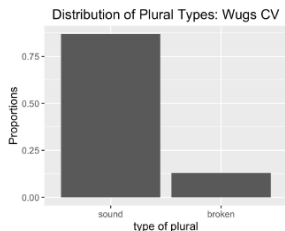
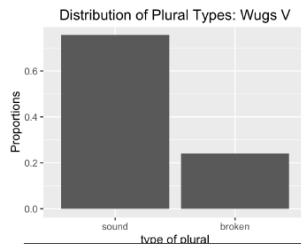
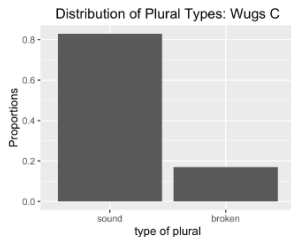
## Qualitative

There is lots a variation in the data:

Nonce	Speaker A	Speaker B	Speaker C	Speaker D
xogol	xgi:gel	xogoliji:t	xogli:t	xogoli
tolluq	tli:laq	tolluqiji:t	tli:qi	tolluqi
żepelp	żepelpiji:t	żpi:pel	żepelpi	
follu	folol	folli	folliji:t	folli:t

# Results

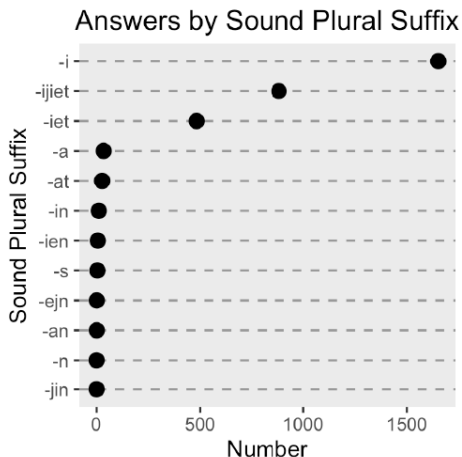
## Sound and broken plurals in nonces and corpus



Corpus results

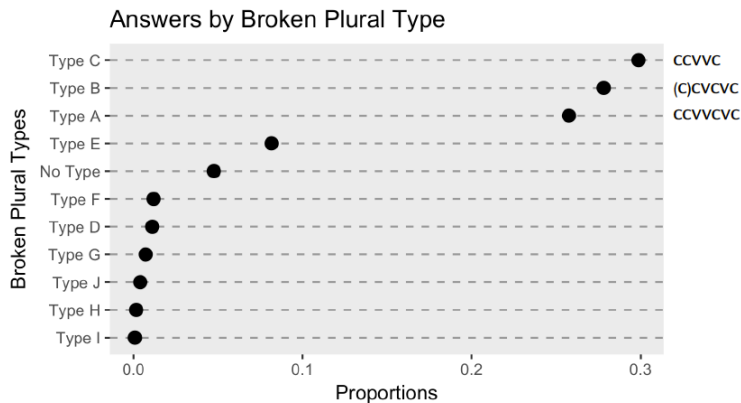
# Results

## Sound plural suffixes



# Results

## Broken plural forms



# Results

## Errors in infrequent forms

Errors frequent		Errors infrequent	
Sound	Broken	Sound	Broken
5 (of 400)	1 (of 400)	14 (of 400)	177 (of 400)
1.3%	0.3%	3.5%	44.3%

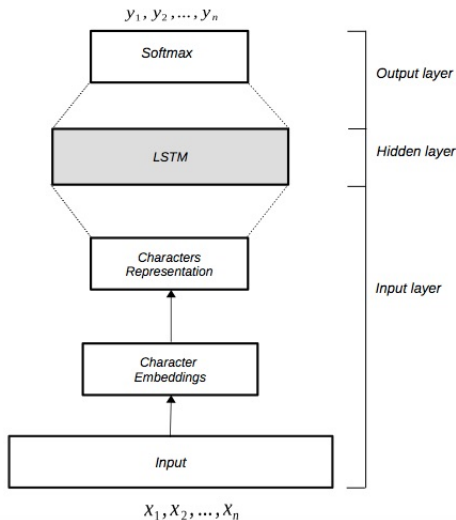
# Models

## Long Short-Term Memory (with the help of Samih Younes)

- Recurrent neural network which we trained to classify Maltese plurals.
  - If there really is no structure at all in the data this should fail.

# Models

## Long Short-Term Memory (with the help of Samih Younes)



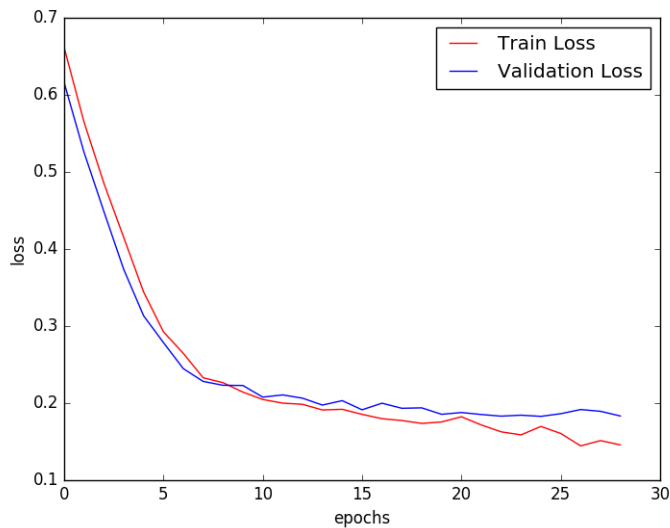
# Models

## Long Short-Term Memory

- data: 2337 word forms (this is based on a version not chacked by ġabra. It contains a few non-nouns.)
- Training: 1869 (broken and sound)
- Validation: 468 (186 broken, 282 sound)

# Models

## Long Short-Term Memory: Learning



# Models

## Long Short-Term Memory: Learning

Label	precision	recall
broken	0.94	0.91
sound	0.94	0.96

# Models

## Long Short-Term Memory and experiment

LSTM	broken	sound
broken	170 (0.36)	16 (0.03)
sound	10 (0.02)	272 (0.58)

Experiment	broken	sound
broken	360 (0.5)	60 (0.08)
sound	0 (0.006)	300 (0.41)

# Long Short-Term Memory and experiment

- There is structure in the data
- The classification is pretty good, broken plurals are underestimated and sound plurals overestimated.

# Minimal Generalization Learner

Albright & Hayes (2003)

- Model that learns by comparing two inflected forms
- The difference between the forms is formulated as rule.
- The differences are generalized over.
  - [dɔg], [dɔgz]:  $\emptyset \rightarrow [z]/[dɔg]_{[plural]}$ .
  - [bæɡ], [bægz]:  $\emptyset \rightarrow [z]/[bæɡ]_{[plural]}$ .
  - generalized:  $\emptyset \rightarrow [z]/X[+voice,-cont]_{[plural]}$ .

# Minimal Generalization Learner

- 2225 corpus pairs as input.
- Tested with 20 new words.

# Minimal Generalization Learner

MGL	broken	sound
broken	3 (0.15)	7 (0.035)
sound	0 (0.0)	10 (0.5)

Experiment	broken	sound
broken	360 (0.5)	60 (0.08)
sound	0 (0.006)	300 (0.41)

# Minimal Generalization Learner

- MGL is essentially a linear model.
- Nevertheless it does well.
- again broken plurals are underestimated and sound plurals a bit overestimated.

# Naive Discriminative Learner

Baayen, Milin, Djurdjević, Hendrix & Marelli (2011)

- Learns associations between *cues* and *outcomes*.
- The cues are singular forms in bigrams
- the outcomes are plural types (sound, broken, bound).
- These associations are weighted.
- we trained the NDL on our corpus.
- We analyzed our nonce words in bigrams and calculated how the NDL learner would classify them.
- The NDL classified 63% the way our participants did.
  - Excluding frequency from the data, the correct classification dropped to 57%.

# Naive Discriminative Learner

In previous models we did not yet model the results of our participants (work in progress). We also had not yet included the category *Bround*. A direct comparison of the models is therefore impossible (fixing this is work in progress).

NDL	Broken	Bround	Sound
Broken	1151 (0.13)	19 (0.002)	757 (0.09)
Bround	38 (0.004)	2 (0.0002)	22 (0.002)
Sound	2131 (0.25)	115 (0.01)	4141 (0.49)

# Naive Discriminative Learner

- NDL does well.
- Model overestimates sound plurals, and is uncertain about  
bround plurals.

# All models

- The performance of the models seems to correlate with the performance of the native speakers (NDL) and the proportions found in the lexicon.
- Especially NDL suggests that phonotactics play an important role in predicting which singular gets what plural.
- It is still difficult to pin down what it is exactly that speakers use as base for the analogies they produce.

# Conclusion

- There is structure in the data.
- Native speakers are able to inflect novel nouns, as sound, broken or bround.
- Several models with very different architectures can learn the plural system relatively successfully.
- NDL learns based on bigrams: dare I say prosodic structure.
- If true, then phonotactics of the singular does indeed determine the plural form.
- There really is no reason to despair.
- (And, as always: much work still needs to be done.)

grazzi ħafna!

## References

- Albright, A. & Hayes, B. (2003). Rules vs. Analogy in English Past Tenses: A Computational/Experimental Study. *Cognition*, 90(2), 119–161.
- Baayen, R. H., Milin, P., Djurdjević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3), 438.
- Borg, A. & Azzopardi-Alexander, M. (1997). *Maltese*. London: Routledge.
- Brincat, J. M. (1996). *Maltese words. An etymological analysis of the Maltese lexicon*. na.
- Mayer, T., Spagnol, M., & Schönhuber, F. (2013). Fixing the broken plural in maltese. *Perspectives on Maltese Linguistics*, 14, 129–158.
- Schembri, T. (2012). *The broken plural in Maltese—A description*. Bochum: Brockmeyer.