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).)
  - ḫabib '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

<table>
<thead>
<tr>
<th>Singular</th>
<th>Plural</th>
<th>Suffix</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>arlogg</td>
<td>arloggį</td>
<td>-i</td>
<td>watch, clock</td>
</tr>
<tr>
<td>omm</td>
<td>ommiji:t</td>
<td>-iji:t</td>
<td>mother</td>
</tr>
<tr>
<td>ṭaddi:m</td>
<td>ṭaddi:ma</td>
<td>-a</td>
<td>worker</td>
</tr>
<tr>
<td>bni:dem</td>
<td>bnedmi:n</td>
<td>-i:n</td>
<td>lazy</td>
</tr>
<tr>
<td>film</td>
<td>films</td>
<td>-s</td>
<td>movie</td>
</tr>
<tr>
<td>saltna</td>
<td>saltni:t</td>
<td>-a:t, -i:t</td>
<td>kingdom</td>
</tr>
</tbody>
</table>

Mayer, Spagnol & Schönhuber (2013)
## Broken plurals

<table>
<thead>
<tr>
<th>Type</th>
<th>Singular</th>
<th>Plural</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>bandiːra</td>
<td>bnaːdar</td>
<td>flag</td>
</tr>
<tr>
<td>B</td>
<td>balla</td>
<td>balal</td>
<td>bundle</td>
</tr>
<tr>
<td>C</td>
<td>borġ</td>
<td>braːġ</td>
<td>heap</td>
</tr>
<tr>
<td>D</td>
<td>xmara</td>
<td>xmajjar</td>
<td>river</td>
</tr>
<tr>
<td>E</td>
<td>xatba</td>
<td>xtaːbi</td>
<td>gate</td>
</tr>
<tr>
<td>F</td>
<td>baħar</td>
<td>iβhra</td>
<td>sea</td>
</tr>
<tr>
<td>G</td>
<td>ġdid</td>
<td>ġdodda</td>
<td>new</td>
</tr>
<tr>
<td>H</td>
<td>għarbi [arbi]</td>
<td>għarab [arap]</td>
<td>Arab</td>
</tr>
<tr>
<td>I</td>
<td>wiċċ</td>
<td>uċuħ</td>
<td>face</td>
</tr>
<tr>
<td>J</td>
<td>għaref [aref] [aref]</td>
<td>għoriːf [oriːf]</td>
<td>wise man</td>
</tr>
<tr>
<td>K</td>
<td>għama [ama]</td>
<td>għomja [omja]</td>
<td>blind person</td>
</tr>
</tbody>
</table>

Schembri (2012)
Several sound plurals for one singular

<table>
<thead>
<tr>
<th>Singular</th>
<th>Plural</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>werqa</td>
<td>werq-at</td>
<td>leaf</td>
</tr>
<tr>
<td>werqa</td>
<td>werq-iːt</td>
<td>leaf</td>
</tr>
</tbody>
</table>
Both sound and broken forms for one singular

<table>
<thead>
<tr>
<th>Singular</th>
<th>Broken plural</th>
<th>Sound plural</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>bandiːra</td>
<td>bnaːdar</td>
<td>bandiːri</td>
<td>flag</td>
</tr>
<tr>
<td>tapit</td>
<td>twapet</td>
<td>tapiti</td>
<td>carpet</td>
</tr>
<tr>
<td>ūnaxix</td>
<td>ūxejjex</td>
<td>ūnaxiːt</td>
<td>vegetables</td>
</tr>
</tbody>
</table>
Some forms seem to have both a suffix and a changed prosody:

<table>
<thead>
<tr>
<th>Singular</th>
<th>Plural</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>bni:dem</td>
<td>bnedm-i:n</td>
<td>lazy</td>
</tr>
<tr>
<td>giddi:b</td>
<td>giddib-in</td>
<td>liar (bround)</td>
</tr>
</tbody>
</table>

Borg & Azzopardi-Alexander (1997)
There is a great deal of variation. It is difficult to pinpoint the rules for sound plurals (Borg & Azzopardi-Alexander, 1997), and broken plurals drive scholars of Maltese to despair: "Dwar il-plural miksur m’hemmx regoli". (There are no rules governing the broken plural. (L-Għaqda Tal-Kittieba Maltin. Cited in: Schembri, 2012)
• 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.
To test these hypotheses we created a corpus and we did a production experiment.
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

Distribution of Plural Types

- Sound
- Broken

proportions

type of plural
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 op 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:

<table>
<thead>
<tr>
<th>Nonce</th>
<th>Speaker A</th>
<th>Speaker B</th>
<th>Speaker C</th>
<th>Speaker D</th>
</tr>
</thead>
<tbody>
<tr>
<td>xogol</td>
<td>xgi:gel</td>
<td>xogoli:j:t</td>
<td>xogli:t</td>
<td>xogoli</td>
</tr>
<tr>
<td>tolluq</td>
<td>tli:laq</td>
<td>tolluqi:j:t</td>
<td>tli:qi</td>
<td>tolluqi</td>
</tr>
<tr>
<td>żepelp</td>
<td>żepelpi:j:t</td>
<td>żpi:pel</td>
<td>żepelpi</td>
<td></td>
</tr>
<tr>
<td>follu</td>
<td>folol</td>
<td>folli</td>
<td>folliji:t</td>
<td>folli:t</td>
</tr>
</tbody>
</table>
Results

Sound and broken plurals in nonces and corpus

![Distribution of Plural Types: Wugs C](image1)

![Distribution of Plural Types: Wugs V](image2)

![Distribution of Plural Types: Wugs CV](image3)

![Distribution of Plural Types](image4)
Results

Sound plural suffixes

Answers by Sound Plural Suffix

- i
- ijet
- iet
- a
- at
- in
- ien
- s
- ejn
- an
- n
- jin
Results

Broken plural forms

Answers by Broken Plural Type

Proportions

Broken Plural Types

Type C
Type B
Type A
Type E
No Type
Type F
Type D
Type G
Type J
Type H
Type I

ccvvc
(c)cvvcvc
ccvvcvc
## Results

### Errors in infrequent forms

<table>
<thead>
<tr>
<th></th>
<th>Errors frequent</th>
<th>Errors infrequent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sound</td>
<td>Broken</td>
</tr>
<tr>
<td></td>
<td>5 (of 400)</td>
<td>1 (of 400)</td>
</tr>
<tr>
<td></td>
<td>1.3%</td>
<td>0.3%</td>
</tr>
</tbody>
</table>
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 checked 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

<table>
<thead>
<tr>
<th>Label</th>
<th>precision</th>
<th>recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>broken</td>
<td>0.94</td>
<td>0.91</td>
</tr>
<tr>
<td>sound</td>
<td>0.94</td>
<td>0.96</td>
</tr>
</tbody>
</table>
# Models

Long Short-Term Memory and experiment

<table>
<thead>
<tr>
<th></th>
<th>broken</th>
<th>sound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LSTM</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>broken</td>
<td>170 (0.36)</td>
<td>16 (0.03)</td>
</tr>
<tr>
<td>sound</td>
<td>10 (0.02)</td>
<td>272 (0.58)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>broken</th>
<th>sound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>broken</td>
<td>360 (0.5)</td>
<td>60 (0.08)</td>
</tr>
<tr>
<td>sound</td>
<td>0 (0.006)</td>
<td>300 (0.41)</td>
</tr>
</tbody>
</table>
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.
  - \([d\,g], [d\,gz]: \emptyset \rightarrow [z]/[d\,g]+_{plural}\].
  - \([b\,g], [b\,gz]: \emptyset \rightarrow [z]/[b\,g]_{plural}\].
  - generalized: \(\emptyset \rightarrow [z]/X[+\text{voice}, -\text{cont}]_{plural}\).
Minimal Generalization Learner

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

<table>
<thead>
<tr>
<th>MGL</th>
<th>broken</th>
<th>sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>broken</td>
<td>3 (0.15)</td>
<td>7 (0.035)</td>
</tr>
<tr>
<td>sound</td>
<td>0 (0.0)</td>
<td>10 (0.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment</th>
<th>broken</th>
<th>sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>broken</td>
<td>360 (0.5)</td>
<td>60 (0.08)</td>
</tr>
<tr>
<td>sound</td>
<td>0 (0.006)</td>
<td>300 (0.41)</td>
</tr>
</tbody>
</table>
Minimal Generalization Learner

- MGL is essentially a linear model.
- Nevertheless is 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, bround).
- 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).

<table>
<thead>
<tr>
<th>NDL</th>
<th>Broken</th>
<th>Bround</th>
<th>Sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broken</td>
<td>1151 (0.13)</td>
<td>19 (0.002)</td>
<td>757 (0.09)</td>
</tr>
<tr>
<td>Bround</td>
<td>38 (0.004)</td>
<td>2 (0.0002)</td>
<td>22 (0.002)</td>
</tr>
<tr>
<td>Sound</td>
<td>2131 (0.25)</td>
<td>115 (0.01)</td>
<td>4141 (0.49)</td>
</tr>
</tbody>
</table>
Naive Discriminative Learner

- NDL does well.
- Model overestimates sound plurals, and is uncertain about bound 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 ġafna!
References


