# Stability and change: computational studies of morphology in contact

Micha Elsner, Ohio State University 2023

## Thanks to my collaborators





Sara K. Court (Ph.D. student, OSU) Andrea D. Sims (Associate Professor, OSU)

This work is partly supported by NSF: "Neural discovery of abstract inflectional structure" (Elsner/Sims BCS 2217554)

#### English in contact

As a major world language, English can't be understood in isolation. For most speakers, English exists in **contact** with other languages.



Adult learners of English as a foreign language...





## English speakers learning other languages...

And members of multilingual communities.



An old-fashioned view was that the bilingual was "two monolinguals in one person" (Grosjean 1989) whose multiple languages rarely interacted.



There is now strong evidence that multiple languages interact via shared mental representations, in both learners and fluent multilinguals.

Matras 2009, Adamou 2021, Kaivapalu and Martin 2007, MacWhinney 2005, among others...

Part 1: Lexical analysis to investigate historical contact

#### Etymology and reanalysis in Maltese

(Court, Elsner, Sims: Intl. Morphology 22; Mediterranean Morphology '22)

#### Part 2: Simulations to understand potential contact outcomes

#### Celtic nouns under English influence

(in submission)

#### How does contact affect languages?

One way to tell is to investigate **historical outcomes** of contact.

Maltese is a Semitic language descended from **Arabic**, but with heavy influence from **Italian** and some **English**.



## "Broken" (non-concatenative) plurals

Many Maltese nouns form plurals by changing the way their vowels are arranged.

This process originals in Arabic, but applies to words of multiple origins in Maltese.

| Singular | Plural   | Gloss           | Allomorph  |
|----------|----------|-----------------|------------|
| fardal   | fradal   | 'apron'         | CCVVCVC    |
| birra    | birer    | 'beer'          | (C)CVCVC   |
| kbir     | kbar     | 'big'           | CCVVC      |
| ftira    | ftajjar  | 'type of bread' | CCVjjVC    |
| bitha    | btieħi   | 'yard'          | CCVVCV     |
| sider    | isdra    | 'chest'         | VCCCV      |
| marid    | morda    | 'sick person'   | CVCCV      |
| ghodda   | ghodod   | 'tool'          | (għ)VCVC   |
| elf      | eluf     | 'thousand'      | VCVC       |
| gharef   | ghorrief | 'wise man'      | CVCCVVC(V) |
| ghama    | ghomja   | 'blind person'  | (għ)VCCV   |

Table 6: Broken plural allomorphs in Maltese, from Nieder et al. (2021a)

## 'Hybrid' or 'stratal' models of the lexicon



Partly-nativized words (grammar I)

Native words (grammar II)

foreign affixes or "elsewhere" class

perhaps restricted set of affixes

loosely based on Ziani 2020 for loans in Arabic; Spagnol 2011, Borg and Gatt 2017 in Maltese

#### Concatenative plurals are also a complex set

Maltese has multiple plural suffixes of different origins.

Deciding which suffix goes with which word is not necessarily simple!

|  | Singular | Plural   | Gloss      | Allomorph  |
|--|----------|----------|------------|------------|
|  | karta    | karti    | 'paper'    | - <i>i</i> |
|  | omm      | ommijiet | 'mother'   | -ijiet     |
|  | rixa     | rixiet   | 'feather'  | -iet       |
|  | giddieb  | giddieba | 'liar'     | - <i>a</i> |
|  | meħlus   | meħlusin | 'freed'    | -in        |
|  | kuxin    | kuxins   | 'cushions' | -S         |
|  | triq     | triqat   | 'street'   | -at        |
|  | sid      | sidien   | 'owner'    | -ien       |
|  | baħri    | baħrin   | 'sailor'   | -n         |
|  | ħati     | ħatjin   | 'guilty'   | -jin       |
|  | spalla   | spallejn | 'shoulder' | -ejn       |
|  | sieq     | saqajn   | 'foot'     | -ajn       |
|  | qiegħ    | qiegħan  | 'bottom'   | -an        |

Table 5: Sound plural allomorphs in Maltese, from Nieder et al. (2021a)

#### Analogical models of the lexicon

sriep?

kbir ~ kbar Common words remain stable because learners have opportunities to observe the correct plural.

Unfamiliar words are modeled with reference to existing ones, which compete to provide alternatives.



See Fertig 2017 for an overview of analogical models

#### Predictions?

How a word sounds is informative about its etymology, and vice versa.

Both accounts predict both of these should matter.

#### Predictions?

How a word sounds is informative about its etymology, and vice versa.

Both accounts predict both of these should matter.

Stratal accounts suggest:

Easier to divide concatenative from non-concatenative (across strata) than to pick the specific process (within stratum).

Surface form mostly redundant once etymology is known.

#### Predictions?

How a word sounds is informative about its etymology, and vice versa.

Both accounts predict both of these should matter.

Analogical accounts suggest:

No high-level split between concatenative and non-concatenative.

Surface form highly informative despite knowing etymology.

#### Methods

We use information theory to quantify the predictive power of our various factors.

Model learns to predict from a set of "training" words We measure its uncertainty on unseen words

LSTM neural network

Following Williams et al. 2020 "Predicting declension class from form and meaning"

| Given:  | Predict:                            |
|---|-------------------------------------|
| Surface form  | Concatenative or non-concatenative? |
| Surface form and etymology                          | Concatenative or non-concatenative? |
| Surface form  | Specific plural type                |
| Surface form and etymology                          | Specific plural type                |
| (*we include gender as a covariate in all analyses) |                                     |

total uncertainty about concatenation

.81 bits of total uncertainty

total uncertainty about concatenation

.81 bits of total uncertainty

etymology and surface form redundantly contribute 6% of this amount tal uncertainty about concatenation

.81 bits of total uncertainty

etymology and surface form redundantly contribute 6% of this amount

etymology contributes up to 13%



.81 bits of total uncertainty

etymology and surface form redundantly contribute 6% of this amount

etymology contributes up to 13%

surface form contributes up to 21%

inty about concatenation

total uncertainty about plural type

.81 bits of total uncertainty

2.65 bits of total uncertainty



.81 bits of total uncertainty

2.65 bits of total uncertainty

etymology and surface form redundantly contribute 15% of this amount

etymology contributes up to 22%

surface form contributes up to 42%

#### Conclusions

Both etymology and surface form contribute non-redundantly to Maltese plural prediction.

Surface form is **more** informative than etymology.

Fine-grained information about the word is **more informative** about the specific plural type than whether the plural is concatenative:

**No evidence** for a high-level split!

#### History is the outcome, but what's the process?

Maltese came about through extensive multilingual contact, but that contact happened in the past.

We can also observe contact reshaping systems in the present.

## Scottish Gaelic language shift (Dorian 1978)

Nancy Dorian studied a community in which Gaelic was being replaced by English.

Multiple plural markers remain even among heritage speakers (the youngest, non-fluent generation), but some types of plural are almost entirely lost.

In particular, **suffixes** seem to survive better than **non-suffixes**.



|   | o.f.s. |     | y.f.s. |      | s.s. |     |
|---|--------|-----|--------|------|------|-----|
| FORMATION TYPE                                    | #      | %   | #      | %    | #    | %   |
| /-(ə)xən/ (I) <sup>a</sup>                        | 48     | 18  | 34     | 12.5 | 87   | 19  |
| final mutation (II)                               | 28     | 10  | 24     | 9    | 23   | 5   |
| /-ən/ (I)   | 27     | 10  | 24     | 9    | 124  | 27  |
| vowel alternation + final mutation (VII)          | 24     | 9   | 28     | 10.5 | 11   | 2.5 |
| vowel alternation + /-(V)n(') <sup>b</sup> (VIII) | 21     | 8   | 33     | 12   | 26   | 5.5 |
| /-čɛn/ (I)  | 20     | 7.5 | 25     | 9    | 18   | 4   |
| lengthening of final consonant (XI)               | 19     | 7   | 15     | 5.5  | 5    | 1   |

"[These share] the fact that the device in question (quantity change) plays no role whatever in English pluralization" - Dorian 1978

Is Dorian right that we can attribute these effects to English contact? How could we (begin to) tell?

#### Comparisons

As Jarvis (2000) points out, it would be methodologically ideal to compare across populations for which everything **except** language background was held constant.

For instance, it would be very convenient to compare English-dominant speakers to speakers whose native language was Arabic or Maltese, for whom non-suffix plurals are common.

But this is rarely possible! Exposure, social attitudes, and learning materials usually all differ.

#### Can simulations tell us about possible outcomes?

Model learns to predict from training words and language tags

We measure its uncertainty on unseen words

transformer neural network





model from Elsner and Court 2022; see also Elsner 2020

Compare simulated learners

Due to data shortage, we use Irish rather than Scottish Gaelic.

We compare:

"Monolingual" (Irish data only)

"L1 English" (28k English nouns, limited Irish)

"L1 Maltese" (28k Maltese nouns, limited Irish)



#### Irish noun plurals

| Suffixal al         | lomorphs: 58%   |      |
|---------------------|-----------------|------|
| ola $\sim$ olaí     | 'oil'           | 21.4 |
| $muc \sim muca$     | 'pig'           | 7.3  |
| inne $\sim$ inní    | 'guts'          | 5.7  |
| rud ~ rudaí         | 'thing'         | 4.8  |
| góraí ~ góraithe    | 'goal'          | 3.9  |
| idé $\sim$ idéanna  | 'idea'          | 3.3  |
| jib $\sim$ jibeanna | 'jib'           | 3.3  |
| lao $\sim$ laonna   | 'calf'          | 1.7  |
| others              |                 | 7    |
| Stem allo           | morphs: 35%     |      |
| bac $\sim$ baic     | 'barrier'       | 22.1 |
| sách $\sim$ sáigh   | 'well-fed pers' | 6.0  |
| fear $\sim$ fir     | 'man'           | 2.1  |
| Síneach ~ Sínigh    | 'Chinese pers'  | 2.0  |
| $mac \sim mic$      | 'son'           | 1.9  |
| others              |                 | 1    |
| Suffix p            | lus stem: 6%    |      |
| cúil ~ cúlacha      | 'corner'        | 1    |
| others              |                 | 5    |

See Carnie 2008, Stenson 2019 for details

#### Hypotheses

- Monolingual models will lag behind with least Irish exposure (boring technical reasons: models perform poorly with little data)
- L1 English model will have an early advantage with suffixes (following Dorian and other language learning studies)
- 3. L1 Maltese model will have an early advantage with non-suffixes
- 4. Models will eventually converge given enough Irish (effects of language transfer are strongest for least-fluent speakers; MacWhinney 2005, among others)











#### Some additional results

L1 English and Maltese models don't just produce **more** suffixal and non-suffixal plurals (respectively)...

But are **better** at producing the correct type.

We conjecture it has learned to attend more to cues at word ends or within stems.

The model produces unmarked (zero) plurals, which Dorian also observes.

#### Conclusions

Computational models can be used as simulated language learners: they undergo **facilitation** and **interference** based on language contact, just like humans.

This makes them useful for hypothesis generation in situations where controlled studies are hard to run.

English L1 influence is consistent with the results Dorian reports; these asymmetries in which patterns are lost are caused by the (suffixal) nature of English, not just by limited exposure to Irish.

#### We favor analogical models of contact

Evidence against models in which learners maintain multiple grammars or coherent lexical strata for the different languages they speak.

Rather, individual words within the system are flexibly assigned to plural types based on cues shared across the linguistic system.

Thus, effects we find or predict are often strongly lexicalized, and predictable from surface word form.

English does not exist in a vacuum. When we imagine English speakers, we should consider what other languages they know, and how contact influences flow in both directions.





Thank you!

#### Matching human ratings of "wug" words



2021 SigMorphon shared task

#### Semitic Lexemes

#### Non-Semitic Lexemes







