Is inflectional irregularity dysfunctional to human processing?

Claudia Marzi, Marcello Ferro, Vito Pirrelli

Institute for Computational Linguistics – National Research Council of Italy

{name.surname}@ilc.cnr.it

Research rationale

The graded complexity of an inflectional system is the resulting equilibrium state of a number of conflicting processing requirements and adaptive responses to task-dependent pressures.

Two principles appear to govern inflectional complexity:

- ✓ **discriminative**: inflected variants must be able to mark the entire space of paradigmatic contrast
- ✓ **implicational**: patterns of variation are interdependent in ways that novel forms can be predicted from known ones.

TEMPORAL SELF-ORGANISING MAPS, a recurrent variant of Kohonen's SOMs, implement a **discriminative/implicational model of word learning**. Tested on cross-linguistic evidence, they shed light on the conflicting processing requirements that any inflectional system must meet, helping us understand the relevance of the contrast between regular and irregular inflection from a processing-oriented functional perspective.

This computational evidence offers novel insights into the **self-organisation of complex inflection systems and the functional architecture of the human word processor**.

Method

TSOMs (*Ferro et al. 2011 ; Pirrelli et al. 2015 ; Marzi et al. 2016*) are trained on a multilingual sample of comparable written inflected verb forms, with no morpho-syntactic and morpho-semantic information.

For each language, and for all languages together, we monitor two types of network response:

- ✓ dynamic levels of per-node activation in a word recognition task
- ✓ accuracy of prediction of the up-coming letter given the preceding context, in the same task

The evidence is analysed with generalised regression models focusing on:

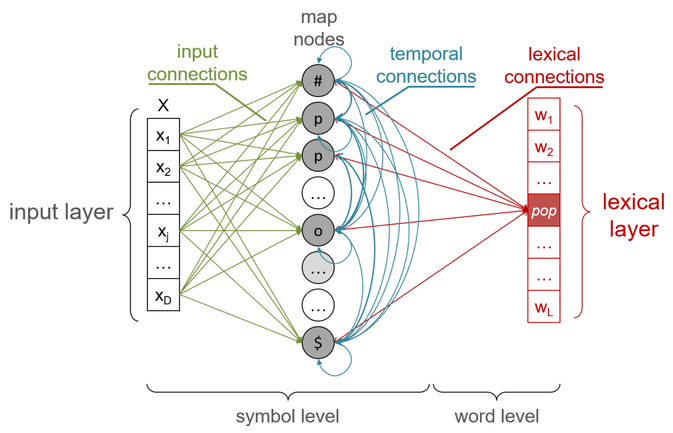
- ✓ interaction effects of **per-node activation** & **per-letter prediction** with regular vs. irregular inflection
- ✓ correlation of these processing effects with perception of emergent morphological structure

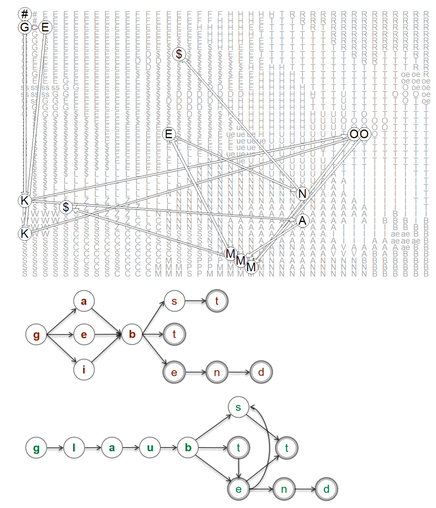
DATA & modelling

We selected 50 high-frequency verb sub-paradigms in MODERN STANDARD ARABIC, ENGLISH, GERMAN, MODERN GREEK, ITALIAN, SPANISH.

To control for implicative relations, a comparable set of max 15 forms for each language was focused on:

* English, German, Italian, Spanish: 6 present and 6 past tense forms, infinitive, gerund/present participle and past participle
* Modern Greek : 6 present, 6 past tense + 3S simple future
* Modern Standard Arabic : 7 imperfective, 7 perfective forms (1S, 2MS, 3MS, 3FS, 1P, 2MP, 3MP) for





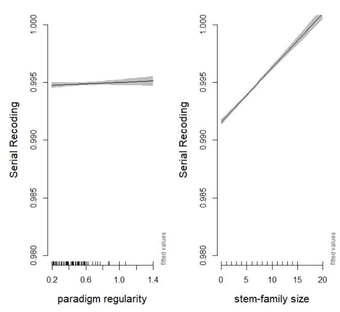
node activation

we monitor the level of activation of the **node maximally responding to the current input letter** in a word recognition task.

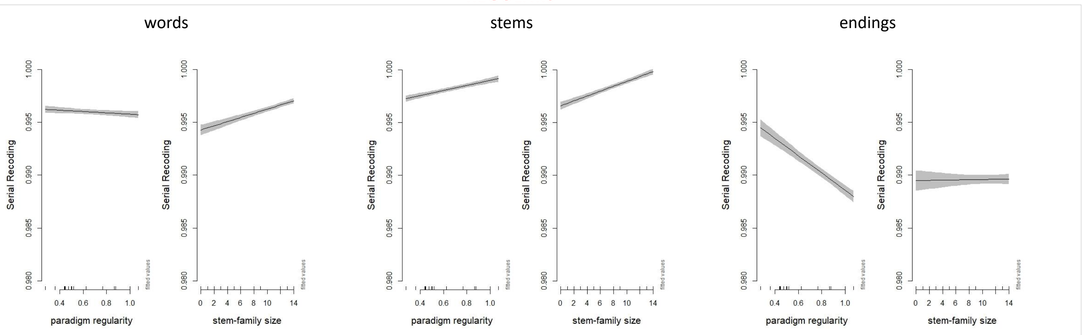
The measure taps into the map's overall level of activation given the observed input, to reflect the paradigmatic level of inflectional self-organisation, where sequences that belong to bigger lexical families tend to fire map nodes more strongly.

Node activation is a reasonable proxy of how easy and familiar an input word is for the map to respond to, thereby modelling a parallel processing task (e.g. visual lexical decision) where wordlikeness or global support matters more than sequential well-formedness.

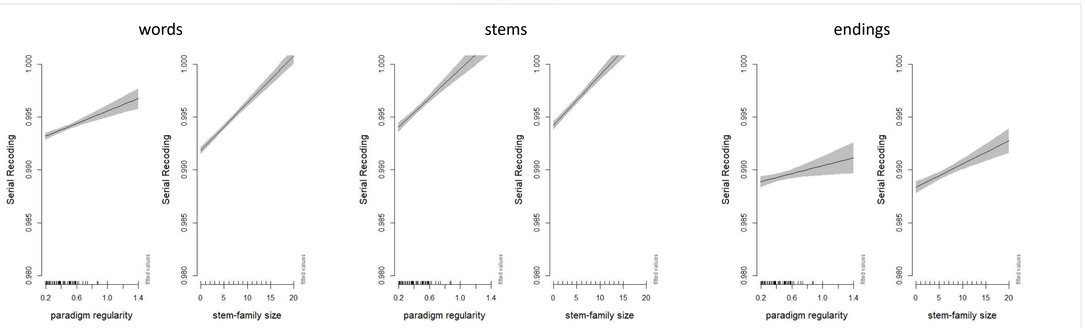
ALL VERB FORMS



REGULARS



IRREGULARS



letter prediction

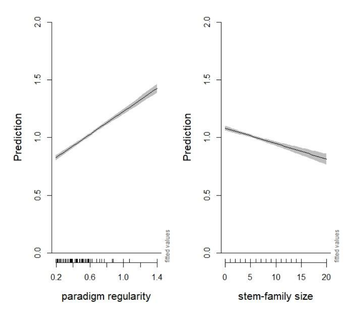
we monitor how well a TSOM can **predict the upcoming input letter** in a word recognition task.

The measure taps into the map's expectation for upcoming symbols at time *t*, through forward temporal connections emanating from the winning node at time *t-1*.

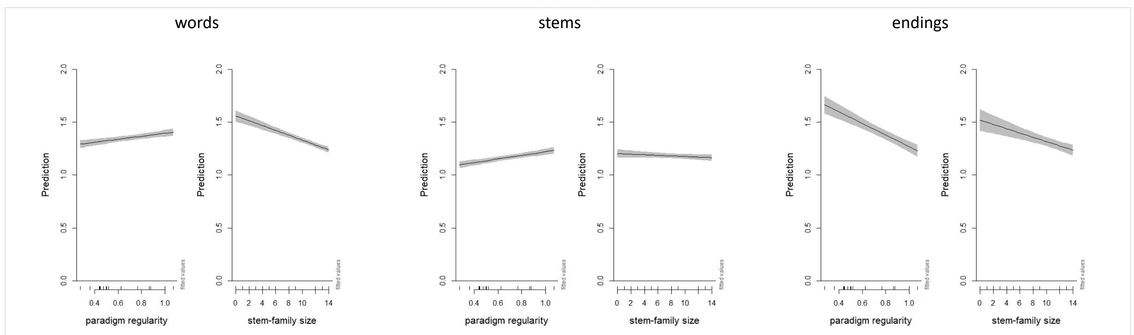
The measure reflects the syntagmatic level of inflectional self-organisation, where the dynamic, conditional probability of a letter sequence is a function of how discriminative and predictive the input sequence is.

Per-letter prediction is a reasonable proxy of how promptly the map is responding to a serial processing task (e.g. acoustic word recognition).

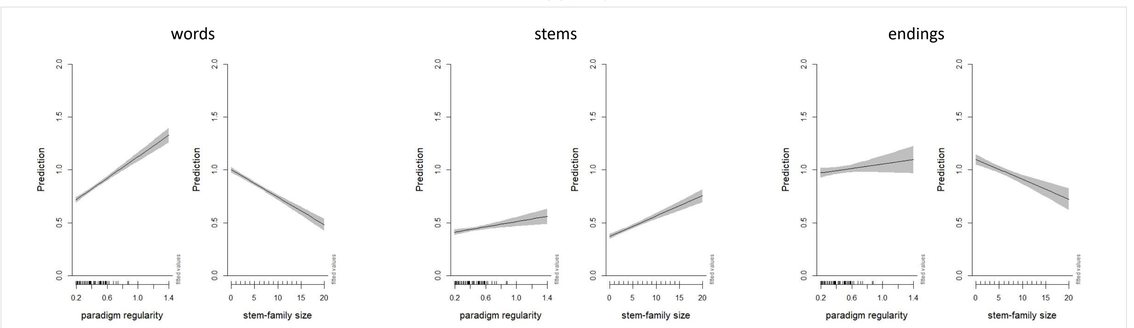
ALL VERB FORMS



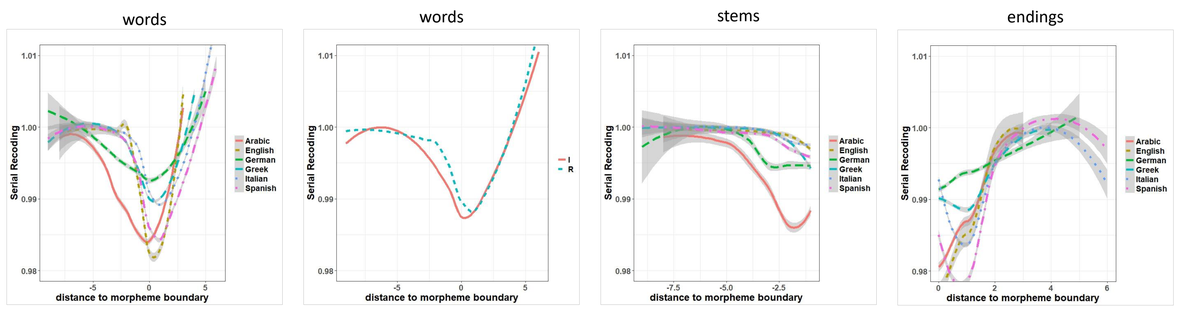
REGULARS



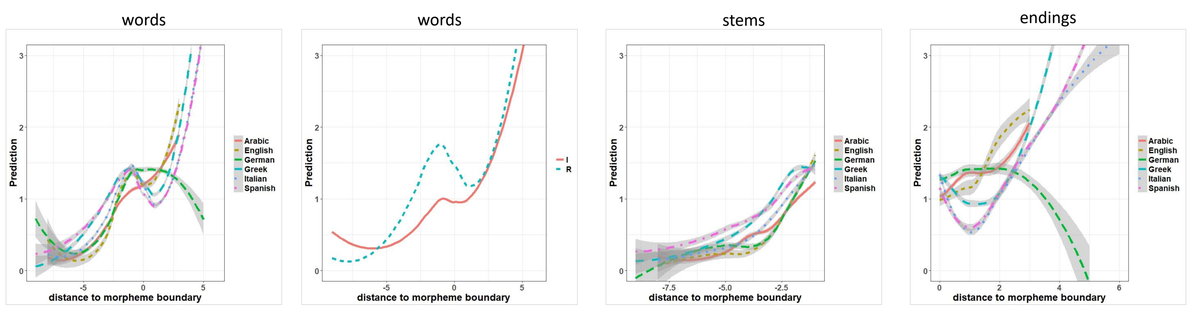
IRREGULARS



non linear effects



non linear effects



processing effects

Irregularity induces a more holistic processing strategy, with reduced perception of structure:

- ✓ the strategy is functional to word processing, when the word point (Uniqueness Point) is reached where a stem is uniquely distinguished from its allomorphs.
- ✓ from that point onwards, the processing pace of irregulars speeds up and compensates for the early disadvantage.

structure sensitivity

TSOM perception of **morphological structure** interacts with inflectional transparency and regularity:

- ✓ sublexical constituents are **more salient** when they remain unchanged across contexts.
- ✓ structural discontinuity increases with the number of contexts where constituents are found.

Main Findings

- ✓ discriminative/implicational learning is **accurate** → 99,62%-99,94% accuracy scores for all language sets
- ✓ morphological irregularity is not dysfunctional →it responds to a **maximally discriminative** function in word processing
- ✓ irregularly inflected forms are typically isolated and processed holistically → irregularity strongly correlates with **token frequency** and **lack of perception** of sublexical structure
- ✓ irregulars have **fewer family members** → irregulars are **more difficult** to acquire
- ✓ regularly inflected forms benefit from repeated patterns of intra-paradigmatic formal redundancy → they are sensitive to family size (or type frequency) effects and effects of structural discontinuity
- ✓ regulars resonate in large paradigm families → regulars are easier to learn implicationally
- ✓ although there is a clear correlation between letter prediction and node activation in TSMOs, the two measures are not mutually implied → a highly predicted node can show a modest level of activation if the sequence the node is activated by is relatively rare or isolated
- ✓ average processing effects show little variance, both cross-linguistically and between regulars and irregulars → variation in processing effects is dynamic and highly non linear