# Evaluating Inflectional Complexity Crosslinguistically: a Processing Perspective

**Claudia Marzi[1], Marcello Ferro[1], Ouafae Nahli[1], Patrizia Belik[2], Stavros Bompolas[3], Vito Pirrelli[1]**
[1]Institute for Computational Linguistics-CNR, [2]Universitat Politècnica de Valencia, [3]University of Patras
patrizia_belik@yahoo.com, stavros.bompolas@gmail.com
{claudia.marzi, marcello.ferro, ouafae.nahli, vito.pirrelli}@ilc.cnr.it

**Abstract**

The paper provides a cognitively motivated method for evaluating the inflectional complexity of a language, based on a sample of "raw" inflected word forms processed and learned by a recurrent self-organising neural network with fixed parameter setting. Training items contain no information about either morphological content or structure. This makes the proposed method independent of both meta-linguistic issues (e.g. format and expressive power of descriptive rules, manual or automated segmentation of input forms, number of inflectional classes etc.) and language-specific typological aspects (e.g. word-based, stem-based or template-based morphology). Results are illustrated by contrasting Arabic, English, German, Greek, Italian and Spanish.

**Keywords:** paradigm-based morphology, inflectional complexity, prediction-based processing, recurrent self-organising networks.

## 1. Introduction

There is little doubt that some languages are inflectionally more complex than others. Everybody would agree with the intuitive statement that the English conjugation system is simpler than the German system, and that the latter is, in turn, simpler than the verb system of Modern Standard Arabic. However, the naïve view is faced with two apparent paradoxes. When linguists try to pinpoint the source of this complexity, the task is far more elusive than expected, and goes well beyond a purely descriptive notion of diversity in the battery of realisational means (e.g. number of different affixes, number of cells in the corresponding paradigms, amount of stem allormophy etc.) provided by each system. Besides, there seems to be a poor correlation between our intuitive notion of morphological complexity and actual evidence of the pace of acquisition of more or less complex inflectional systems in child language. In some cases, apparently simpler inflectional markers may take more time to be acquired than formally more complex and articulated ones. What looks like a prohibitively difficult learning task in the light of the complexity and uncertainty of the inference steps required for mastering it, may turn out to be relatively unproblematic for human speakers. In the present paper we entertain a usage-oriented, cognitively motivated approach to issues of morphological complexity, based on a neurobiologically inspired model of word processing and learning, and explore its theoretical and computational implications.

## 2. Background

Assessing and understanding the comparative complexity of the inflectional system of a language relative to a functionally-equivalent system of another language remains an open question, which has animated much of the contemporary debate on the nature of word knowledge and its connection with issues of word usage (Ackerman and Malouf, 2013; Bane, 2008; Bearman et al., 2015; Juola, 1998; Moscoso del Prado Martín et al., 2004). In a crosslinguistic perspective, the way morphosyntactic features are contextually realised through processes of word inflection probably represents the widest dimension of grammatical crosslinguistic variation, somewhat belittling universal invariances along other dimensions (Evans and Levinson, 2009).

Descriptive linguists have often approached the issue of comparative inflectional complexity by providing comprehensive catalogues of the morphological markers and patterns in a given language or languages (Bickel and Nichols, 2005; McWorther, 2001; Shosted, 2006). Accordingly, the complexity of an inflectional system is measured by simply enumerating the number of category values instantiated in the system (e.g. person, number or tense features) and the range of available markers for their realisation: the bigger the number, the more difficult the resulting system. The notion of Enumerative Complexity (or E-complexity) is however dubious (Ackerman and Malouf, 2013). Suppose we have two hypothetical inflectional systems, each with two categories only (say, singular and plural) and three different endings for each category: A, B, C for singular, and D, E and F for plural. In the first system, paradigms are found to present three possible pairs of endings only: <A, D>, <B, E>, <C, F> (corresponding to three different inflection classes). In the second system, any combination is attested. Clearly, the latter system would be more difficult to learn than the former, as it makes it harder to infer the plural form of a word from its singular form. Nonetheless, both systems present the same degree of E-complexity.

Of late, less combinatorial approaches to morphological description have played down the role of E-complexity in inflection. These approaches, generally referred to as "paradigm-based", or "word-based", or "abstractive" grammatical frameworks, examine the systemic organisation of underlying patterns of surface variation, to conceive of an inflectional system as a network of implicative relations holding between fully-inflected forms (Blevins, 2003; Blevins, 2016; Burzio, 1998; Bybee, 1995; Bybee and McClelland, 2005; Matthews, 1991; Pirrelli and Battista, 2000). Implicative relations allow novel forms to be predicted and inferred on the basis of known forms, thereby making it easier for a human speaker to process, retain and access them. Not only do implicative relations shed light on the way children come to master the inflectional system of their mother tongue, but they also constrain systems of word shapes, providing a limit on the range of E-complexity that languages can afford.

A number of information theoretic approaches have been proposed to model this view in terms of Kolmogorov complexity (Kolmogorov, 1965) and Shannon entropy (Shannon, 1948). The idea behind Kolmogorov complexity is to measure a dataset of inflected forms as the shortest possible grammar needed to describe them. This however leads to a definition of morphological complexity heavily dependent on the grammar formalism adopted (Bane, 2008; Walther and Sagot, 2011). Ackerman and Malouf (2013) use Shannon's information entropy to quantify prediction of an inflected form as a paradigm-based change in the speaker's uncertainty. They conjecture that inflectional systems tend to minimise the average conditional entropy of predicting each form in a paradigm on the basis of any other form of the same paradigm (Low Conditional Entropy Conjecture or LCEC). This is measured by looking at the distribution of inflectional markers across inflection classes in the morphological system of a language. Although LCEC proves to be able to capture a substantial part of the inferential complexity within paradigms, it presupposes a segmentation of inflected forms into stems and affixes, while ignoring implicative relations holding between stem allomorphs. Use of principal parts can remedy this in a principled way (Finkel and Stump, 2007, among others). However, while entropy measures can provide extremely valuable insights into the organisation of static, synchronic paradigms, there are crucial complementary questions about how such patterns are processed and learned which remain unaddressed. In what follows, we will focus on these important issues from a neuro-computational perspective. In particular, we are interested in evaluating the net effect of the complexity of an inflectional system on the processing behaviour of a recurrent neural network, excluding the role of word token frequency effects on prediction-driven processing (Pickering and Garrod, 2013). To factor out frequency effects, we ran simulations on uniformly distributed inflectional data. Our work can hence be understood as a purely morphological evaluation of complexity, based on lexical rather than corpus data. Since uniform distributions increase the entropy of a system, our results define some sort of upper bounds for inflectional complexity: if all factors (including frequency) are taken into account, the effects we observe here will likely be more prone to potentially confounding factors. This is in the spirit of information-theoretic work on paradigm-based morphology, as well as 'discriminative learning' research in animal behaviour and language learning (Rescorla and Wagner, 1972; Ramscar and Yarlett, 2007; Ramscar and Dye, 2011), and justifies our choice of a specific type of recurrent neural network, namely a Temporal Self-Organising Map (Ferro et al., 2011; Marzi and Pirrelli, 2015; Pirrelli et al., 2015; Marzi et al., 2016) as a workbench for simulating pardigm-based effects. Ultimately, it is intended to bridge the gap between an algorithmic/mathematical understanding of processing-based morphological complexity (Balling and Baayen, 2008; Balling and Baayen, 2012), and the neurobiological (or implementational) level of Marr's hierarchy (Marr, 1982).

## 3. Method and data

According to Dressler and colleagues (Bittner et al., 2003), European languages can be arranged along an inflectional complexity continuum, ranging from a more inflecting-fusional type (left) to a more isolating type (right):

*Lithuanian→Greek→Russian→Croatian→Italian→ Spanish→German→Dutch→French→English.*

Somewhat paradoxically, developmental evidence provides an indication that inflectional contrasts in prototypically inflecting verb systems are reported to be acquired at an earlier stage than inflectional contrasts in more isolating verb systems.[1]

Here, we would like to investigate the related question about how degrees of inflectional complexity/regularity affect word processing strategies. For this purpose, we analyse the performance of recurrent self-organising neural networks learning a few languages in the typological continuum above: namely, English, German, Greek, Italian and Spanish. To broaden our typological data, Standard Modern Arabic was added to the range of tested languages. For each language we sampled the 50 top-frequency verb paradigms found in a few reference resources: CELEX (Baayen et al., 1995) for German and English; the Paisà Corpus (Lyding et al., 2014) for Italian; the European Spanish Subcorpus of the Spanish TenTen Corpus (www.sketchengine.co.uk); the SUBTLEX-GR corpus (Dimitropoulou et al., 2010) for Modern Greek; the Penn Arabic Treebank (Maamouri et al., 2004). To control paradigm implicative relations, we selected a comparable set of 15 paradigm cells (14 cells for Arabic).[2] The sample contains a shared set of 6 present and 6 past tense forms for English, German, Greek, Italian and Spanish. Infinitive, gerund/present participle and past participle forms were added for English, German, Italian and Spanish, whereas 3 singular forms of the simple future were included for Modern Greek. The Arabic set contains 7 imperfective and 7 perfective forms, including 1S, 2MS, 3MS, 3FS, 1P, 2MP, 3MP cells. Only inflected "raw" forms from the selected cells were included for training a recurrent neural network, with no additional morphological information. Each language-specific dataset is administered to a Temporal Self-Organising Map (hereafter *TSOM,* see section 3.2. for more details) for 100 epochs. In one epoch, all word forms are randomly input to the map five times, and each training session was repetated five times with results averaged over repetitions to control random variabil-

[1]For example, Noccetti (2003) reports that the transition from pre- to proto-morphology in Italian verb acquisition has an early onset at Brown's stage II, with mean length of utterance 2 (Brown, 1973), in contrast with the comparative late emergence of the third-person singular marker *-s* in the acquisition of the English present tense.

[2]The full set of data, for each language, is available at http://www.comphyslab.it/redirect/?id=lrec2018_data

ity. TSOM parameters are identically initialised across the 6 languages, with the only exception of available memory nodes (Table 1).[3]

## 3.1. The data

The selected paradigm cells for the target languages offer evidence of graded levels of morphological (ir-)regularities. Greek, Italian, Spanish and German present highly inflecting conjugation systems, with extensive stem allomorphy, exhibiting varying degrees of (ir)regularity. Inflecting processes include prefixation, suffixation, vowel alternation, infixation and suppletion. Arabic stem formation is based on the interspersion of discontinuous consonantal roots and variable vowel patterns. English offers the by far simplest inflectional system, with extensive syncretism and a rather dichotomous subdivision of paradigms between regular and irregular ones.

For all test languages except Modern Greek, word forms are orthographically transcribed, and administered to the network one symbol at a time as raw letter strings (starting with the start-of-word symbol *'#'* and ending with the end-of-word symbol *'$'*), with no information about their morphological structure. To account for the complex interaction between morphologically-conditioned and phonologically-conditioned stem allomorphy in Greek conjugation (Ralli, 2005; Ralli, 2006), Greek word forms are transcribed phonologically, and input one segment at a time. Once more, no information about morphological structure is input. To assess the network sensitivity to morphological structure and the processing behaviour of the map across morpheme boundaries (see section 4.), after training, word forms in all test languages were segmented morphologically according to a prefix-stem-suffix schema: e.g. Greek *e-krin-a* 'I judged', German *ge-dach-t* 'thought' (past participle), Arabic *ya-ktub-u* 'he writes'. Stem allomorphs within a single paradigm (whether morphologically/phonologically predictable or not) are segmented as whole units, with no explicit indication of either the root or the alternating pattern: e.g. Arabic *katab-a* 'he wrote' vs. *ya-ktub-u* 'he writes'. Only purely suffixal stem formation is segmented: e.g. Greek *aɣapi-s-a* 'I loved', Italian *perd-ut-o* 'lost' (past participle).

## 3.2. Recurrent self-organising neural networks

TSOMs are recurrent self-organising networks consisting of two-dimensional grids of artificial memory/processing nodes that learn to dynamically memorise input strings as chains of maximally-responding processing nodes (Best Matching Units, or BMUs), whose level of sensitivity to input symbols in context is a continuous function of their distributional regularities in training (Ferro et al., 2011; Marzi and Pirrelli, 2015; Pirrelli et al., 2015; Marzi et al., 2016). In a TSOM, each processing node has two layers of synaptic connectivity: an input layer, connecting each node to the current input stimulus (i.e. orthographic or phonological symbols), and a (re-entrant) temporal layer, connecting each node to all other nodes. Every time a symbol is presented to the input layer, activation propagates to all map nodes through input and temporal connections, and the most highly activated node (BMU) is calculated (see Figure 1). Given the BMU at time $t$, the temporal layer encodes the expectation of the current BMU for the node to be activated at time $t+1$. The strength of the connection between consecutively activated BMUs is trained through the following principles of discriminative learning: given the input bigram *ab*, the connection strength between the BMU that get mostly activated for *a* at time *t* and the BMU for *b* at time $t+1$ will:
(i) increase if *a* often precedes *b* in training (entrenchment),
(ii) decrease if *b* is often preceded by a symbol other than *a* (competition).

The complex interaction between entrenchment and competition in a TSOM accounts for important dynamic effects of self-organisation of stored words (Marzi et al., 2014; Marzi et al., 2016). In particular, at a sublexical level, systematically recurrent patterns tend to recruit context-sensitive specialised (and stronger) chains of BMUs. If the bigram *'ab'* is repeatedly input to the TSOM, the map tends to develop a specialised BMU('*b*') for '*b*' in '*ab*' and a highly-weighted outward connection from BMU('*a*') to BMU('*b*'), reflecting a strong expectation of BMU('*a*') for a prospective BMU('*b*'). In detail, during training, weights on both connectivity layers are adjusted in an experience-dependent fashion: after an initial period of random variability, where nodes activate chaotically, a map gradually develops more and more specialised sequence of BMUs for word forms - or sub-lexical chains - that are functionally dependent on the frequency distribution and the amount of formal redundancy in the training data. On the one hand, specialised inter-node connectivity makes BMUs less confusable and more salient, as they receive stronger support through temporal connections than any other node. On the other hand, less specialised and more blended BMUs are densely and less strongly connected with many others, to meet the input of more words. When a TSOM is trained

| *language* | form length min/max | paradigms reg./irreg. | word types/ training size | TSOM nodes |
|---|---|---|---|---|
| Arabic | 4/11 | 18/28 | 560/601 | 40x40 |
| English | 2/11 | 20/30 | 208/750 | 35x35 |
| German | 3/11 | 16/34 | 504/750 | 40x40 |
| Greek | 2/13 | 37/13 | 744/750 | 42x42 |
| Italian | 2/12 | 23/27 | 748/750 | 42x42 |
| Spanish | 2/15 | 23/27 | 715/750 | 40x40 |

Table 1: Language training sets. Form length is measured by the number of orthographic/phonetic symbols. In the Italian sample, we encoded the orthographic accent as a separate character (e.g. *è* = e' ). Differences between word types and cardinality of the training set are due to syncretism (particularly extensive in English). Paradigm defectiveness explains the smaller cardinality of the Arabic training set.

[3] For the sake of data comparability, the number of memory nodes for each language was decided empirically to control for cross-linguistic differences in cardinality and length of word types (see Table 1). For all trained languages, the percentage of used nodes among all available nodes ranges between 31% and 35%.

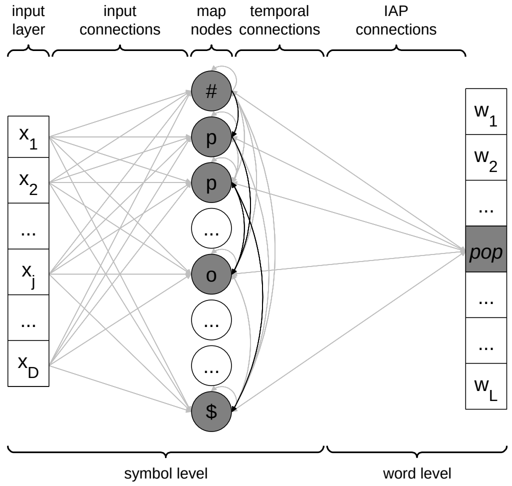


Figure 1: Functional architecture of a Temporal Self-Organising Map (TSOM). Each input word form is presented by a unique time-series of symbols, which are administered one at a time.

on higly redundant input data such as verb paradigms, specialisation and blending may interact. By inputting all verb forms with a uniform token distribution, we factor out the effect of frequency and focus our analysis on the effect of formal redundancy only. Thus, due to the prediction-driven bias of the temporal layer of re-entrant connections, strong expectations over upcoming input symbols account for successful serial word processing, with processing accuracy being a function of how confident the TSOM is about the position of the current symbol in the input string.

These dynamics make it possible to test the behaviour of a TSOM on specific lexical tasks: *word recall* and *serial word processing*. For each time series of input symbols (i.e. each word form), the processing response of the map is represented by the synchronic activation pattern of all the BMUs that most highly get activated for that input sequence. Thus, the task of *word recall* tests how accurately a map can retrieve the input word from its synchronic activation pattern, namely how accurately the activation nodes of the map can encode information about the timing of the input symbols that make up the word. Accuracy in recall verifies that, for each input form, activation propagation (i.e. sequential activation) of nodes within each synchronic pattern correctly activates the BMUs associated with the symbols of each word. Scores are given in Table 2, showing very high accuracy and remarkably cross-linguistic similarity.

Conversely, *serial word processing* can be monitored by evaluating the ability of a map to predict an incrementally presented input word. Proceduraly, by presenting one symbol at a time on the input layer, a TSOM is prompted to complete the current input string by anticipating the upcoming BMU to be activated. Anticipation/prediction scores across input words are calculated by incrementally assigning each correctly anticipated symbol in the input form a 1-point score, i.e. the anticipation score of the preceding symbol incremented by 1. Otherwise, for unpredicted symbols the score is 0. The more input symbols are anticipated,

| *language* | recall % | sd % |
|---|---|---|
| Arabic | 99.93 | 0.16 |
| English | 99.62 | 0.86 |
| German | 99.76 | 0.18 |
| Greek | 99.84 | 0.06 |
| Italian | 99.79 | 0.15 |
| Spanish | 99.94 | 0.13 |

Table 2: For each language, percentage values of correctly recalled word types and standard deviations are given, averaged over 5 map instances.

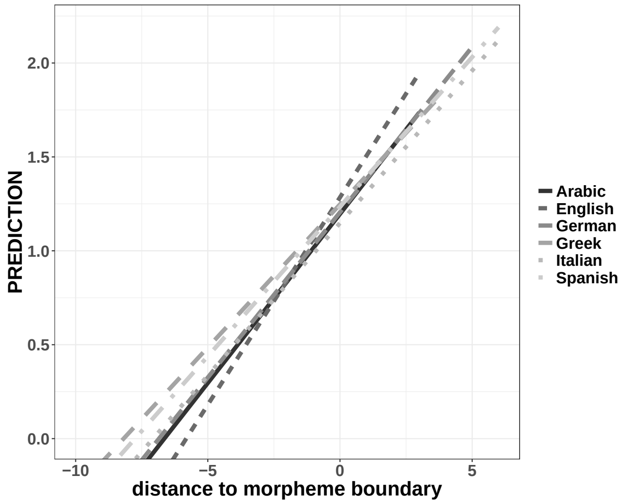


Figure 2: Marginal plot of interaction effects between language and distance to morpheme boundary, in an LMER model fitting the number of symbols predicted by TOMs. Fixed effects: languages, distance to morpheme boundary. Random effects: TSOM instances, paradigms, word forms.

the easier the prediction of that word. Results will be given and discuss in the ensuing section (4.).

## 4. Results and discussion

All results in this section are analysed with linear models for mixed effects (LMERs). For all models/languages, we treated TSOM instances, verb paradigms and word forms as random effects. In particular, we show how inflectional systems of different complexity (independent variables) affect TSOM processing, by focusing on symbol prediction rate as a dependent variable.

Figure 2 plots, for each language, the rate of symbol prediction in serial word processing. It should be appreciated that Arabic, German, Italian and Spanish exhibit remarkably similar trends, with not significantly different slopes (p-values >.05). Only Greek and English present significantly different slopes (p-values <.001), with Greek forms being the hardest to process (lower slope), and English forms the easiest ones (higher slope).

To evaluate the impact of formal transparency on processing, the effect of regularity is fitted in a second LMER model where languages are considered as random effects. Across our selected languages, verb forms in regular paradigms are systematically more predictable (p-value


Figure 3: Marginal plot of interaction effects between categorical (ir-)regularity and distance to morpheme boundary, in an LMER model fitting the number of symbols predicted by TOMs. Fixed effects: irregulars (I) vs. regulars (R), distance to morpheme boundary. Random effects: languages, TSOM instances, paradigms, word forms.

<.001) than forms in irregular ones, as shown by the marginal plot in Figure 3.

To investigate in more detail the impact of inflectional complexity on processing, we fitted an LMER of symbol prediction for each language, with classes of morphological regularity (regulars vs. irregulars) and morphological structure (stem vs. suffix) as fixed effects (Figure 4).

The marginal plots in Figure 4 better show a clear serial processing effect of the distance of an input symbol to the stem-ending boundary, over and above the length of the input string. Unsurprisingly, Italian and Spanish show a very similar behaviour, with irregular forms exhibiting fusional effects that blur the boundary between stem and inflectional endings, and comparable (but not identical) number of stem allomorphs (Boyé and Cabredo Hofherr, 2006; Pirrelli, 2000). Remarkably, both German and Greek exhibit systematic (albeit not always predictable) processes of stem formation, followed by a fairly homogenous pool of inflectional endings. As a result, in both languages, the base stem (or present stem) is often followed by a highly embedded and unpredictable sequence of symbols which account for the negative slopes in the corresponding segments. In Arabic imperfective forms, prefixation is used to convey person features. This makes selection of inflectional endings fairly predictable, given the stem. Finally, in our pool of languages, English offers the by far simplest inflectional system, with extensive syncretism and a rather dichotomous subdivision of paradigms between regular and irregular ones.

Slopes are also modulated by degrees of regularity/transparency of the stem. Discontinuous patterns of morphological structure are often found in irregular paradigms of concatenative languages (e.g. English *drink/drunk*, German *finden/fanden*), and are systematically attested in non-concatenative morphologies (e.g. Arabic *kataba/yaktubu*). It is well known in the literature on serial alignment that discontinuous patterns are more difficult to be processed and tracked down (Hahn and Bailey, 2005). In the present context, stem allomorphs are less predictable since their "uniqueness point", i.e. the point at which they can be distinguished from all neighbouring allomorphs, are normally delayed, slowing down processing (Balling and Baayen, 2012). Other things being equal,[4] the order of magnitude of this competing effect is a function of the number of stem allomorphs: the more they are, the more confusable the input stem is. Conversely, in regular paradigms the same stem shows up systematically in all cells. Hence the stem suffers from no intra-paradigmatic competition. These factors provide, on average, a net processing advantage of stems in regular paradigms, as confirmed by the significant difference in prediction rate between stems of regular vs. irregular paradigms in all languages (Figure 4). However, the clear advantage in stem processing is somewhat compensated by the difference in the prediction rate on suffixes. In German, Greek, Italian and Spanish suffixes in irregulars are predicted significantly more easily than suffixes in regular forms, as shown by the steeper segments in the positive

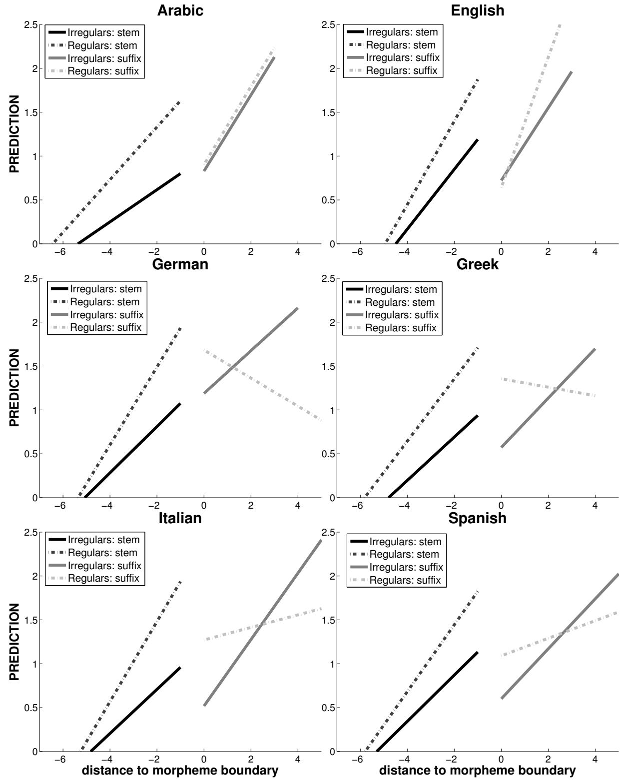


Figure 4: For each language, marginal plots of interaction effects between morphological (ir-)regularity and distance to morpheme boundary, in LMER models fitting the number of symbols predicted by TOMs for stem and suffix. Fixed effects: regularity (dashed lines) vs. irregularity (solid lines), distance to morpheme boundary, stem and suffix as separate patterns, suffix length. Random effects: TSOM instances, paradigms, word forms.

[4]The effect is modulated by other factors we are not controlling here: i.e. the formal similarity between the input stem and its intra-paradigmatic competitors, the entropy of the paradigm, the lexical neighbourhood of the word form.

$x$ range of Figure 4. Besides, for all languages, there is a deeper drop in prediction rate at the stem-suffix boundary (for $x = 0$ as the first symbol of the suffix) in regular forms. In fact, stem allomorphs typically select only a subset of paradigm cells. Hence they can be followed by fewer inflectional endings than regular stems are. This reduces processing uncertainty, by constraining the range of possible continuations at the stem-suffix boundary of irregularly inflected forms. As a result, irregulars tend to blur the TSOM sensitivity to the verb morphological structure, favouring a somewhat more holistic processing strategy.

Results and statistical significance are confirmed when we consider a more fine-grained meausure for inflectional complexity based on a gradient of morphological regularity, which takes into account the number of stem alternants of a given paradigm.[5] It represents a graded - and continuous - meausure of paradigmatic (ir-)regularity that considers, for each inflected form, the number of stem-sharing forms (or stem family size), instead of a dichotomous and formal classification of paradigms (regulars vs. irregulars). Thus, given the number of inflected form-types for each paradigms, the average stem family size correlates better with the non-categorical idea of inflectional complexity.

## 5. Concluding remarks

Our evidence is in line with Low Conditional Entropy Conjecture (Ackerman and Malouf, 2013). The processing cost of considerably different inflectional systems appears to oscillate within a fairly limited range of variation, whose upper bound and lower bound are marked, in our language sample, by Modern Greek and English respectively. All other conjugations present no statistically significant differences in the processing overhead they require, in spite of their typological diversity, which is nonetheless reflected by the different processing profiles exhibited by sublexical constituents in the different languages.

In a functional perspective, this evidence can be interpreted as the result of a balancing act between two potentially competing communicative requirements: (i) a recognition-driven tendency for a maximally contrastive system; and (ii) a production-driven bias for a maximally generalisable inflection system, where, for each paradigm, all forms in the paradigm can possibly be deduced from any one of its forms.

This interpretation is also compatible with another clear pattern shown by our data. In each of our sample languages, the difference between the processing cost of forms in irregular paradigms compared with the processing cost of forms in irregular paradigms shows an interesting structure-sensitive profile. The higher processing cost of irregular stems is compensated by a lower cost in processing the inflectional endings selected by irregular stems. Once more, these structural effects tend to reduce processing costs at the level of the whole word, making the inflectional system as functional as possible from an information theoretic perspective. In recognising that scale effects play an important role in the processing behaviour of our model at the word level, and that constrains on word processing are likely to obtain universally, we also highlight the fundamental communicative role of words as optimal-sized units for describing general functional tendencies in language, and for studying language as a complex information system.

Inflectional complexity is multifactorial and dynamic. Its variability can be observed and measured on many counts: number and types of stem allomorphs, number and types of inflectional affixes, transparency/compositionality effects, stem-stem predictability, stem-affix predictability, affix-affix predictability, intra-paradigmatic and inter-paradigmatic frequency distributions etc. In this paper, we investigated inflectional complexity by controlling a number of interacting factors through language-specific training regimes, on which we ran a psycho-linguistically plausible computer model of inflection learning. In this way, we could understand more of factor interaction through a quantitative analysis of the way the performance of our system is affected across different training regimes. Methodologically, it allows for much more flexible and controlled test/analysis protocols than those commonly used with human subjects in experimental psycholinguistics.

In addition, understanding more of the real cognitive hurdles a human learner has to face in the process of effectively acquiring an inflectional system of average complexity may also shed some light on optimal practices for language teaching.

[5]This graded notion takes into accout exceptional alternating stems in otherwise regular paradigms (e.g. Italian *aprire/aperto* and Spanish *abrir/abierto*, "open" infinitive/"opened" past participle). At the same time, it captures the difference between partially irregular paradigms and radically idiosyncratic ones.

## 6. Bibliographical References

Ackerman, F. and Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language*, 89(3):429–464.

Baayen, H. R., Piepenbrock, P., and Gulikers, L., (1995). *The CELEX Lexical Database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Balling, L. W. and Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from danish. *Language and Cognitive processes*, 23(7-8):1159–1190.

Balling, L. W. and Baayen, R. H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*, 125(1):80–106.

Bane, M. (2008). Quantifying and measuring morphological complexity. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, pages 69–76.

Matthew Bearman, et al., editors. (2015). *Understanding and Measuring Morphological Complexity*. Oxford University Press.

Bickel, B. and Nichols, J. (2005). Inflectional synthesis of the verb. In David Gil Martin Haspelmath, Matthew S. Dryer et al., editors, *The World Atlas of Language Structures*, pages 94–97. Oxford University Press.

Dagmar Bittner, et al., editors. (2003). *Development of Verb Inflection in First Language Acquisition: a cross-linguistic perspective*. Mouton de Gruyter, Berlin.

Blevins, J. P. (2003). Stems and paradigms. *Language*, 79(2):737–767.

Blevins, J. P. (2016). *Word and Paradigm Morphology*. Oxford University Press, Oxford.

Boyé, G. and Cabredo Hofherr, P. (2006). The structure of allomorphy in spanish verb inflection. *Cuadernos de Lingüística*, 13:9–24.

Brown, R. (1973). *A first language: the early stages*. George Allen I& Unwin.

Burzio, L. (1998). Multiple correspondence. *Lingua*, 104:79–109.

Bybee, J. and McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22(2-4):381–410.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5):425–455.

Dimitropoulou, M., Duñabeitia, J. A., Avilés, A., Corral, J., and Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behavior: the case of greek. *Frontiers in Psychology*, 1(218):1–12.

Evans, N. and Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, (32):429–92.

Ferro, M., Marzi, C., and Pirrelli, V. (2011). A self-organizing model of word storage and processing: implications for morphology learning. *Lingue e Linguaggio*, 10(2):209–226.

Finkel, R. and Stump, G. (2007). Principal parts and morphological typology. *Morphology*, 17:39–75.

Hahn, U. and Bailey, T. M. (2005). What makes words sound similar? *Cognition*, 97:227–267.

Juola, P. (1998). Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics*, 3(5):206–213.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of information transmission*, 1(1):1–7.

Lyding, V., Stemle, E., Borghetti, C., Brunello, M., Castagnoli, S., Dell' Orletta, F., Dittmann, H., Lenci, A., and Pirrelli, V. (2014). The paisá corpus of italian web texts. Proceedings of the 9th Web as Corpus Workshop (WaC-9)@ EACL 2014, pages 36–43. Association for Computational Linguistics.

Maamouri, M., Bies, A., Buckwalter, T., and Mekki, W. (2004). The penn arabic treebank: Building a large-scale annotated arabic corpus. In *NEMLAR conference on Arabic language resources and tools*, volume 27, pages 466–467.

Marr, D. (1982). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman and Company.

Marzi, C. and Pirrelli, V. (2015). A neuro-computational approach to understanding the mental lexicon. *Journal of Cognitive Science*, 16(4):493–535.

Marzi, C., Ferro, M., and Pirrelli, V. (2014). Morphological structure through lexical parsability. *Lingue e Linguaggio*, 13(2):263–290.

Marzi, C., Ferro, M., Cardillo, F. A., and Pirrelli, V. (2016). Effects of frequency and regularity in an integrative model of word storage and processing. *Italian Journal of Linguistics*, 28(1):79–114.

Matthews, P. H. (1991). *Morphology*. Cambridge University Press, Cambridge.

McWorther, J. (2001). The world's simplest grammars are creole grammars. *Linguistic Typology*, (5):125–166.

Moscoso del Prado Martín, F., Kostić, A., and Baayen., R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94(1):1–18.

Noccetti, S. (2003). Acquisition of verb morphology in italian: A case study. In Dagmar Bittner, et al., editors, *Development of Verb Inflection in First Language Acquisition: a cross-linguistic perspective*, pages 351–378. Mouton de Gruyter, Berlin.

Pickering, M. J. and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36:329–392.

Pirrelli, V. and Battista, M. (2000). The paradigmatic dimension of stem allomorphy in italian verb inflection. *Italian Journal of Linguistics*, 12(2):307–380.

Pirrelli, V., Ferro, M., and Marzi, C. (2015). Computational complexity of abstractive morphology. In Matthew Baerman, et al., editors, *Understanding and Measuring Morphological Complexity*, pages 141–166. Oxford University Press, Oxford.

Pirrelli, V. (2000). *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Istituti Editoriali e Poligrafici Internazionali, Pisa.

Ralli, A. (2005). *Morfologia [Morphology]*. Patakis, Athens.

Ralli, A., (2006). *On the role of allomorphy in inflectional morphology: evidence from dialectal variation*, pages 123–152. Polimetrica, Monza.

Ramscar, M. and Dye, M. (2011). Learning language from the input: Why innate constraints can't explain noun compounding. *Cognitive Psychology*, 62(1):1–40.

Ramscar, M. and Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, 31(6):927–960.

Rescorla, R. A. and Wagner, A. R. (1972). A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In Abraham H Black et al., editors, *Classical conditioning II: Current research and theory*, pages 64–99. Appleton-Century-Crofts, New York.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Shosted, R. (2006). Correlating complexity: a typological approach. *Linguistic Typology*, (10):1–40.

Walther, G. and Sagot, B. (2011). Modélisation et implémentation de phénomènes flexionnels non-canoniques. *Traitement Automatique des Langues*, 2(52):91–122.