

# ADAPTIVE STRATEGIES IN LEXICAL ACQUISITION

---

CLAUDIA MARZI

MARCELLO FERRO

**ABSTRACT:** The emergence of morphological patterns from lexical storage in language acquisition is conditioned by language-specific factors as well as extra-linguistic cognitive capacities. With particular reference to the acquisition of plural markers in German, in a memory-based perspective highlighting interesting theoretical implications for usage-based models, the paper analyses acquisitional strategies by focussing on emergent relations between stored word forms and on dynamic expectation/competition of incoming input. In particular, we outline an adaptive multifactorial account of morphological processing that includes both frequency and formal factors. Our investigation is supported by a computational model of morphology acquisition/processing based on self-organisation memories, where word representations are dynamically recoded as time-series.

**KEYWORDS:** word processing, morphological generalisation, German plurals, self-organising memory.

## 1. INTRODUCTION

One of the most important issues in language acquisition is represented by the emergence of a morphological organisation level in the lexicon, viewed by many scholars as a domain of interface between phonology, morpho-syntax and semantics, resulting from a process of lexical self-organisation (Jackendoff 2002). In an *ABSTRACTIVE* perspective (Blevins 2006), in contrast with a constructive approach to word structure which assumes a redundancy-free lexicon based on roots and affixes as the basic building blocks of morphological competence, self-organisation of fully-stored word forms is a determinant of lexical competence, with adaptive strategies for lexical acquisition and processing relying on emergent morphological patterns.

Focusing on inflectional morphology in particular, the emergence of morphological patterns from lexical storage is conditioned by formal factors such as richness, uniformity and transparency of inflectional paradigms in the input. Due to the typological variety of existing morphologies across the world languages, however, a child is faced with an exceedingly unconstrained space of alternative strategies for morphological marking, ranging from morpheme-based affixation, to position-based templatic structures, to process-based phenomena like reduplication (Bybee 1985; Anderson 1992;

Croft 2001; Stump 2001; Haspelmath 2002). Specifically, from a logical standpoint the dynamic process of morphology acquisition thus requires (i) that morphologically complex word forms are scanned to search for their recurrent morphological formatives (morphological segmentation), (ii) that patterns of sub-lexical formatives are related to the morpho-lexical and morpho-syntactic content of their embedding word forms, and (iii) that generalisation strategies must occur to understand and produce novel word forms.

As finding distributional regularities is a fundamental challenge in language acquisition, we focus here on the role of the emergence of common and recurrent patterns in language data. Schematically, the task of inducing morphological knowledge from word forms must consist of two main sub-tasks: (i) finding structure in word forms, and (ii) grouping word forms on the basis of the amount of shared structure.

According to USAGE-BASED (UB) models, these mechanisms of acquisition are highly sensitive to input properties such as type and token frequency, and semantic and phonological consistency (Lieven & Tomasello 2008; Tomasello 2003). Foremost, lexical frequency is a determinant of entrenchment of emergent schemas. Token frequency is defined by how often an individual word form is presented in the input. Forms with high token frequency are known to leave deeply entrenched memory traces in the mental lexicon (Alegre & Gordon 1999; Baayen et al. 2007), and are accessed more quickly and produced accurately. Type frequency counts the word members of a given inflectional class (e.g. the number of different stems that get the inflectional suffix *-ed* in the English past participle). Inflectional classes with high type frequency tend to be most productive and easy to use. This is because, in accord with UB approaches to lexical knowledge, morphological schemas are acquired through a process of generalisation across numerous stored items in the speaker's lexicon. The productivity of a schema like  $[\text{VERB } [-ed]]_{\text{past.tense}}$  defines a way of yielding past tense forms from English verb stems. Although schemas are a function of lexical development, they abstract away from specific lexical entries, and depend on a critical mass of known forms to develop generalisations that apply to novel forms.

According to Bybee's model (1995, 2002), both type and token frequency are important determinants of speaker's increasing accuracy with inflection morphology. However, type frequency is acknowledged to be more relevant to the issue of causing a critical mass of input schemas to emerge and become productive (Bybee 2008). Moreover the dynamic between type and token frequency has a bearing on the rate of acquisition of schematised patterns. Individual word forms which are repeatedly shown in the input get memorised and eventually perceived as wholes, with little sharing of redundantly specified sub-lexical structure with other morphologically-related word forms. Although this appears to favour acquisition at

the level of individual items, holistically-stored forms do not seem to play a distinctive role in the overall organisation of lexical knowledge. This has two important consequences on morphological processing: (i) high-frequency words do not take part in morphological schemas, and (ii) they do not support acquisition of other morphologically-related forms.

The idea that frequency distributions can shape morphological productivity and play a fundamental role in lexical acquisition has been criticised on several grounds (Marcus et al. 1995; Pinker & Ullman 2002). In particular, the existence of so-called minority defaults, i.e. general productive inflectional patterns which are in fact neither supported by numerous lexical types nor possibly entrenched through high-frequency items, has been shown to pose a serious challenge to UB models (Plunkett & Nakisa 1997). German noun plurals in *-s* are probably the best known example of a minority default, as they apply across the board, with no phonological constraints on their bases, but nonetheless appear to operate when other, overwhelmingly more frequent plural patterns fail to apply (Marcus et al. 1995; Sonnenstuhl & Huth 2002; Kauschke et al. 2011).

In the present contribution, we intend to face this challenge by exploring the role of lexical frequency in memory entrenchment and lexical organisation. In particular, we would like to establish an explanatory connection between mechanisms of lexical storage, which are universally known to be governed by frequency factors, and mechanisms of morphological processing and generalisation, which are commonly understood as requiring rule-like formal tools. We show that morphological schemas can be accounted for as the result of the long-term entrenchment of neural circuits (chains of time-stamped memory nodes) that are repeatedly being activated by a memory map in the process of recoding input words into a two-dimensional lexical layer.

The investigation we propose is supported by a computational model of morphology acquisition and processing, based on Self-Organising Maps (SOMs, Kohonen 2001), in a variant of classical SOMs augmented with re-entrant Hebbian connections defined over a Temporal layer (TSOMs), which can encode probabilistic expectations upon incoming stimuli (Koutnik 2007; Ferro et al. 2010; Pirrelli et al. 2011; Ferro et al. 2011; Marzi et al. 2012a; Marzi et al. 2012b). Temporal first-order connections, providing the state of activation of a map at the immediately preceding time step, can be interpreted as encoding the map's probabilistic expectations of up-coming events on the basis of past experience, making room in this way for memorising time series of symbols as activation chains of nodes.

By looking at issues of morphology acquisition from this perspective, we show that usage-based factors such as storage mechanisms, principles of activation, competition processes among concurrently stored lexical forms, type and token frequency distributions conspire in nontrivial ways to

produce processing behaviours that range from full storage to default-like rules, with no need for separate processing mechanisms or segregated modules. Although frequency cannot tell us the whole story, the multi-factorial view of morphological processing that we suggest here goes a long way in accounting for apparently differential effects in word storage and processing, to suggest that the notion of adaptive processing, i.e. the entrenchment of patterns of activation that are repeatedly and successfully used in word recognition, is key to understanding issues of human word knowledge.

In the ensuing sections, in a memory-based perspective we will investigate the developmental trajectory of German noun plurals acquisition as an adaptive multi-factorial process, with the support of computer simulations based on temporal self-organising memories (TSOMs). First, we set out our approach against the debate between one-route and dual-route models of inflection acquisition and processing (Section 2). We then reconsider the case of German plurals from a memory-based perspective, by highlighting its theoretical implications and connections with usage-based models (Section 3). Section 4 is dedicated to an analysis of simulations of German plural marker acquisition with TSOMs. Finally, some general conclusions are drawn (Section 5).

## 2. BEYOND THE SINGLE VS. DUAL ROUTE DEBATE

The so-called PAST-TENSE DEBATE has dominated the theoretical background, on the architecture of morphological lexicon and its relation to grammar, since 1986 (Rumelhart & McClelland 1986; Pinker & Prince 1988).

According to the DUAL-ROUTE approach to word structure (Pinker & Prince 1988; Prasada & Pinker 1993; Pinker & Ullman 2002; Clahsen 1999; Marcus et al. 1995) recognition – and production – of a morphological complex input word involves two interlocked steps: (i) a first preliminary full-form access (the lexical way), and (ii) a second optional morpheme-based access of sub-constituents of the input word, resulting from application of combinatorial rules (the grammatical way). This second step is taken if and only if the first one fails to find any matching access entry; that is the case when a word is not memorised as a full form in the lexicon.

Such a view supports the idea that irregular past-tense forms, as well as all the other morphologically complex irregular word forms, are stored in the lexicon, together with all few high-frequency regular forms.

A radically alternative approach, the mainstream CONNECTIONIST answer to word storage and processing (Rumelhart & McClelland 1986), assumes a single route model which defines a direct correspondence between related word forms, i.e. an input base word form and an inflectionally-related output form. Morphological structure plays no direct role here, and it is an epiphenomenal by-product of the identity mapping between invariant portions of input and output patterns.

Interestingly, these two alternative approaches seem to share two fundamental assumptions: (i) the view that regular inflection is the outcome of a DERIVATIONAL relationship between a lexical base and an inflected form, with the former being preliminarily available, and the latter being produced on-line, and (ii) the idea that both input and output representations are part of the training environment, not the end result of an acquisition process.

Over the past three decades, the psycholinguistic literature has shed novel light over the debate on morphological acquisition and competence. In the perspective of a word-based theory of morphological competence and an approach to the lexicon as a dynamic system, a growing body of empirical findings suggests that the theoretical idea of morphological forms being derived in isolation from their bases is psychologically implausible, and that surface word relations constitute a fundamental domain of morphological competence. A large number of variables have been identified to have an influence on lexical processing, such as for example word length, semantic concreteness, age of acquisition (in case of L2), and word-frequency. In particular, frequency of full forms and size of morphological families are understood to have an influence on lexical processing (Baayen et al. 1997; Taft 1979; Hay 2001; Moscoso et al. 2004a), and surface word relations and paradigmatic structures do play an important role in lexical organisation (Moscoso 2007).

In line with this body of evidence, the dynamic, usage-based perspective we are entertaining here focuses on emergent relations between fully stored word forms and on the dynamic expectation and competition of incoming input forms. Word processing and lexical acquisition are implemented as recoding and storage strategies for time-series of symbols, which rely on both language-specific factors (such as distribution of input forms and patterns of morphological structure) and extra-linguistic cognitive functions (e.g. memory self-organisation, lexical access and recall).

### *2.1 TSOM modelling of memory self-organisation*

A TSOM architecture consists of a grid of topologically-organised memory nodes, representing one layer of neurons in a cortical map, with dedicated sensitivity to time-bound stimuli. Compared to classical Kohonen's Self-Organizing Maps (SOMs; Kohonen 2001), TSOMs are augmented with weighted re-entrant inter-node connections which propagate activation from a given node to the pool of its most likely followers, i.e. those nodes that are most often activated soon after it during (unsupervised) training. Through sheer exposure to time-series of symbolic units (e.g. letters making up the words of a lexicon), neighbouring nodes become increasingly sensitive to recurrent input patterns, i.e. input stimuli which are similar in both encoding and distribution. This is attained through incremental training. At each time tick, one stimulus is input to the map to concurrently activate all nodes. The most highly activated node

(or Best Matching Unit, BMU) gets more sensitive to the current input stimulus and strengthens its pre-synaptic connection to the BMU at the previous time tick. The dynamic is conducive to node competition, specialisation and memory entrenchment. The more often an input sequence is shown to the map, the more likely the sequence is to recruit dedicated/specialised nodes only. This can be interpreted as entrenchment. Conversely, low-frequency sequences tend to activate nodes that are shared by other similar sequences. Shared nodes are shared memory structures, which in turn favour spread of information. After training, since a TSOM in a strict sense does not offer an output representation, but only recodes input patterns on a single layer of map circuitry, the KNOWLEDGE of a trained map is stored in the synaptic weights of its nodes.

In the implementation proposed by the authors (for a more detailed description of the architecture see Ferro et al. 2011; Marzi et al. 2012b), a TSOM is designed to simulate and test models of word storage and processing. Words are represented as temporal symbolic patterns, i.e. strings of acoustic/written symbols that are produced and input one symbol at a time, under the assumption that time paces are sampled discretely upon the event of symbol production/presentation. Time plays a distinctive role in word processing. Past information offers evidence on how to process and store incoming input; whereas generalisation requires the ability to understand unseen forms based on the discovery of recurrent sub-lexical constituents, whose coding is both context-sensitive and fairly independent of specific positional slots (Marzi et al. 2012a, 2012b). In processing an input word, the map activates several BMUs (one for each symbol in the input). For each input word, we call ACTIVATION CHAIN the integrated activation pattern that includes all BMUs triggered by the word. Input words that share some morphological structure may activate partially overlapping activation chains. Due to competition and specialisation, however, high-frequency words, as opposed to low-frequency words, are less likely to activate overlapping chains.

In what follows, we investigate how different frequency distributions of training data may affect the developmental trajectory of a TSOM through the combined interplay of competition/expectation, entrenchment based on cumulative type-frequency effects, and familiarisation processes which condition expectations on incoming input, both known and unknown.

### 3. A MEMORY-BASED REAPPRAISAL OF GERMAN NOUN PLURAL INFLECTION

German inflectional morphology offers very interesting evidence for a reappraisal of a memory-based approach to morphology acquisition, and for analysing the role of type frequency, similarity and structure in the language input (Hahn & Nakisa 2000).

The complexity of German noun plural formation is defined by five different plural patterns, partially combined with an umlaut process. In detail, the suffixes *-e*, *-er* and the absence of a suffix can be combined with vowel changing, while *-(e)n* and *-s* are never combined with an umlauting process.

In a dual-route perspective, German plural morphology has been identified as giving support to the identification of a default pattern – namely the *-s* suffix – all other suffixes and the vowel changing process being taken as irregular patterns (Marcus et al. 1995). Although the *-s* suffix represents only a small percentage of plural nouns, Marcus and colleagues identify this infrequent plural pattern as a default rule not because it represents a REGULAR process in a descriptive sense (i.e. not because it applies to the vast majority of German forms), but because it occurs with different and heterogeneous word patterns thus representing the prototypical example of a memory-independent rule to be applied.

A more nuanced approach has been proposed in an updated dual mechanism model (Bartke et al. 2005), which nonetheless remains anchored to the idea that different, *a priori* determined, processes must be assumed to account for regular and irregular plurals. More challenging for our present concerns is the authors' observation that a unitary model, based on pattern association and frequencies, cannot offer a possible alternative explanation to the overgeneralisation of the *-s* pattern to both novel and nonce words.

The critical assumption here is that one-route pattern associators can generate default behaviour only for a statistically predominant type, whereas dual-route models suffer no such limitation. In fact, as observed by Hahn & Nakisa (2000), the behaviour of dual-route models is entirely determined by the interaction between rules and the associative component. Hence, the question of whether a dual-route system can adequately capture a realistic distribution of German noun plurals is entirely empirical and depends on the actual implementation of the lexical thresholds blocking rule application.

Here, we want to pursue a different line of argumentation and show that the differential behaviour of *-s* plurals with respect to other purportedly non default German plurals can be accounted for in terms of the complex interaction between frequency distributions, which represent a fundamental determinant of memory processes, and other concurrent factors, such as the formal consistency of an inflectional pattern, transparency/iconicity of paradigms and prediction biases induced by the concurrent competition of fully-stored forms in the speakers' mental lexicon.

In line with an emergentist perspective on the interaction between storage and processing, grounded on distributional information and memory-based strategies (MacWhinney 1987; Bates & MacWhinney 1987; Ellis 2002, 2006), TSOMs offer a promising computational framework for modelling morphology acquisition and processing, by focusing on different effects of

entrenchment versus competition, type/token frequency interaction, generalisation by extension both within a class and between classes. Emergent morphological patterns play a very important role in word processing and acquisition, and “when linguistic structure is viewed as emergent from the repeated application of underlying processes, rather than given a priori or by design, then language can be seen as a complex adaptive system” (Bybee 2010: 2).

### *3.1 Generalisation and productivity in a memory-based perspective*

Generalisation of an inflectional class *A* is a function of its productivity, i.e. its capacity to attract lexical members, or lexical types. Commonly, this is possible whenever a few lexical members of *A* share some features: e.g. a root-final sound sequence, a particular gender, a particular semantic class, or a combination of these and possibly yet other features. When a few of them are systematically presented by a subset of the members of an inflectional class, productivity represents a tendency to assign some novel words sharing those features to the same class (Bybee & Hopper 2001). Productivity is thus a function of high type frequency (the class contains many lexical members) and regularity (members systematically share some distinctive features).

Two inflection classes may compete when the same set of features is shared by members of both classes. When this is the case, an inflectional class *A* may possibly attract members of a different inflectional class *B*. For example, the plural form \*LICHTE (i.e. the nonstandard plural alternative, or the wrongly produced \*LICHTEN) instead of the expected LICHTER in German noun inflection, reflects the competition between different plural suffixes triggered by a shared pattern in members of different classes, e.g. the coda -CHT in NACHRICHT, ANSICHT (-en plural class), GESICHT, GEWICHT (-e plural class), and MACHT, NACHT (-e plural class + umlaut). Competition causes -e plurals, or alternatively -en plurals, to intrude into the -er class of plurals, thus co-opting some of its members (e.g. LICHT). The more types share some class-distinctive features the stronger their role in attracting new types. Type frequency thus correlates with the productivity of an inflectional class, although typically type frequency alone is a necessary but not sufficient condition for productivity.

The notion of default productivity represents an apparent case of productivity without regularity. According to dual-route modellers, the German -s class of plural nouns attracts all and only those words which do not fit into other, non-default inflection classes. On the other hand, one-route modellers (Hahn & Nakisa 2000; among others) interpret a default rule as a rule whose conditioning features spread over a large feature space than more specific (subregular) patterns.

Token frequency favours entrenchment of a lexical type, which is no longer perceived as an analogue of other members of its own inflection class.



As a result, the token frequency of a type counteracts productivity of the class to which the type belongs. This is because, in an entrenched type, distinctive features (e.g. a particular ending) are no longer perceived as shared by other types, but are rather part and parcel of the entrenched type.

This is in line with frequency-based (or usage-based) approaches to productivity (Hay & Baayen 2002) and appears to be in good accord with experimental evidence of time latencies in lexical decision task, which are shown to correlate negatively with token frequency, paradigm size and paradigm entropy (Moscoso et al. 2004b).

All this evidence points to the existence of a deep interconnection between memory and processing strategies, suggesting the view that the way humans store and organise full words in their mental lexicon is deeply affected by the way they perceive them during processing. In TSOMs, lexical memory traces reflect repeatedly successful strategies of time-bound word recoding, which are accumulated and applied over again whenever possible, in word recognition and production. This means that stored word patterns affect, in turn, processing, by building expectations on up-coming stimuli. Deeply entrenched, isolated patterns facilitate recognition of familiar words, but fail to contribute to recoding novel words (e.g. the highly frequent singular and plural forms MANN-MÄNNER, with a frequency of respectively 2883 and 989 tokens do not contribute to recode the plural form RÄDER, attested in our dataset only in the singular form RAD). On the other hand, interconnected patterns favour extension and generalisation of known information to unknown cases (this is the case of coherent islands of reliability represented by, for example, the plural forms ABTEILUNGEN, BEDIENTUNGEN, EINRICHTUNGEN, HOFFNUNGEN, WOHNUNGEN, which contribute to generalise the non-attested plural form BEDEUTUNGEN). This shows that memory and processing are really two different dynamic aspects of a unitary process which unfolds through lexical acquisition.

In this perspective, we intend to investigate a few properties of the German noun plural system to focus on the dynamic relation between regularity, productivity and competition of inflection patterns through computer simulations of type-frequency and token-frequency effects.

#### 4. GERMAN NOUN PLURAL EXPERIMENTS

Two-hundred-and-fifty top frequency ranked noun pairs – both singular and plural – were selected from CELEX (Baayen et al. 1995) (with a mean token frequency of 441.09 and standard deviation of 625.87). Accuracy in recoding, recalling and generalisation is tested to monitor the dynamic relation of type/token frequency effects and their sensitivity to and impact on the overall dynamic of a TSOM and its topological organisation.

The task of RECODING consists in testing the accuracy of the map's

activation on input word forms, both trained and untrained. For each symbol shown to the TSOM, we verify if the map recodes the symbol correctly by activating the appropriate labelled node (BMU; for detailed description of temporal layer plasticity and long-term potentiation mechanisms see Pirrelli et al. 2011; Marzi et al. 2012a). When all the symbols of an input word form are recoded correctly, then the input word is recoded accurately. It must be noted that recoding accuracy requires faithful memory traces of currently input symbol, and is at the same time a function of how well a current input symbol can be expected/predicted on the basis of past symbols. In other words, the recoding task is one measure of expectation/prediction and short-term storage.

The *RECALLING* task is intended to verify how well the map can reconstruct the correct sequence of symbols making up a word, by reading this information off the integrated activation pattern triggered by showing the word to the map one symbol at a time. Elsewhere, we argued that the ability of a map to recall an input word is the result of the dynamic interaction of memory trace sustainment and long-term storage of lexical information (Marzi et al. 2012b, for a detailed outline on the architecture and lexical recall task). It can be easily understood how this second task gives an additional important measure of prediction and long-term storage in a TSOM.

The ratio of the two measures of recoding and recall accuracy portrays the overall dynamic of a TSOM and its adaptive self-organisation to paradigmatic relations induced by input data. Namely, sensitivity to symbol identity (as opposed to symbol timing) makes TSOMs more able to capture morphological structure, whereas sensitivity to symbol timing is a condition *sine qua non* for memorising and recalling stored traces.

Finally, a *GENERALISATION* task provides a measure of the ability of a TSOM to recode/recall novel words. This is based on how well the map can predict sequences that are not seen in training, and requires forward spreading of activation through post-synaptic temporal connections (Marzi et al. 2012b). In this perspective, generalisation is a by-product of adapting memory self-organisation strategies for word recoding, as it relies on storage, processing and paradigmatic extension. More importantly for our present concerns, generalisation requires finding distributional regularities and recurrent patterns in language data.

In this connection, German noun plurals represent a challenging test for evaluating the different roles of type vs. token frequency, competition vs. entrenchment processing, and the role of similarity in language input structure.

#### *4.1 Materials, methods and results*

In a first experiment, all the 500 selected word forms were presented 5 times each to a 45×45 nodes map, for one hundred learning epochs. We ran 5

repetitions of the same experiment with unmodified training parameters, and averaged result scores (recode accuracy 100%, recall accuracy 96.48% –  $\sigma$  0.8%).

In a second experiment, we trained a 35×35 nodes map with 300 of the selected word forms, presented with their actual token frequencies, in 5 repetitions of unmodified training parameters (recode accuracy 100%, recall accuracy 94.78% –  $\sigma$  2.5%). Details of the distribution of plural classes in both datasets are given in Table 1.

PLURAL CLASS	DATASET	
	500 TYPES	300 TYPES
∅	3%	2%
∅ + Umlaut	1%	1%
-e	32%	17%
-e + Umlaut	13%	18%
-(e)n	39%	44%
-er	3%	3%
-er + Umlaut	4%	6%
-s	5%	9%
	100%	100%

TABLE 1. PLURAL CLASSES NOUNS PERCENTAGE FOR BOTH DATASETS

TSOMs are known to be sensitive to global effects of paradigm organisation: different morphological families, exhibiting different levels of formal redundancy, tend to have different effects on the way a map organises its topology (Marzi et al. 2012a; Marzi et al. 2012b). To show the global effects of paradigm organisation, we first analysed the overall topology of the resulting maps, and compared it with other similar experiments based on the paradigmatic organisation of the German verb system (Marzi et al. 2012c), where different types of related intra- and inter-paradigmatic verb families induce a strongly paradigm-driven co-organisation which appears to facilitate paradigmatic extension and generalisation. Verb topological clusters tend to group distributionally-related nodes, i.e. nodes that activate in response to similarly-distributed stimuli. As a result, the same input symbol may recruit topologically scattered nodes, each of which is sensitive to a particular positional instantiation of that symbol in the input. On the other hand, the overall topological organisation of TSOMs trained on German nouns appears to be more sensitive to symbol identity than symbol distribution, with each letters activating a topologically-connected cluster of neighbouring nodes. This is shown in Figure 1, which illustrates the topological organisation of a TSOM trained on the 250 singular/plural noun pairs. The map exhibits dedicated sensitivity to input symbol identity, with a notable exception: umlauted vowels appear to activate time-bound nodes, which cluster with nodes encoding symbols that occur interchangeably.

This difference in the topological organisation of nouns and verbs in German inflection can be interpreted as a side-effect of the different paradigmatic organisation of the two systems. Verbs are organised into larger paradigm families (more inflected forms related to the same lexical exponent). Therefore, the verb map appears to be more sensitive to shared and repeated patterns in the input data. On the other hand, nouns are organised into smaller paradigms, and this smaller amount of redundancy in the input seems to account for a poorer sensitivity of the noun map to contextually-based occurrences of individual input symbols.



FIGURE 1. TOPOLOGICAL ORGANISATION OF A 45×45 NODES TSOM TRAINED ON GERMAN NOUNS AT FINAL LEARNING EPOCH. EACH MAP NODE IS LABELLED WITH THE INPUT SYMBOL THAT MOST HIGHLY ACTIVATES THE NODE

To provide a quantitative estimate of this effect on topological organisation, we compared a TSOM trained on verb forms (750 types from 50 paradigms containing 15 inflected forms each) with a TSOM trained on noun forms. For each map, we measured its topological connectedness, i.e. its propensity to associate each different symbol with a single connected cluster of nodes labelled with that symbol. This is computed as the ratio between the number of different symbols in the input, and the number of topologically non-connected clusters of map nodes with the same label. The score equals 1 if the map develops one connected cluster per symbol and goes down to 0 if it exhibits a large number of highly fragmented clusters with the same label.

In Figure 2, we plotted values of map connectedness over learning epochs for nouns (thick line), showing a clustering strategy based on the

conservation, throughout learning, of a one-symbol-one-cluster organisation. On the contrary, verbal topology (Figure 2, thin line) tends to adapt itself, during learning, to the emergence of shared both intra- and inter-paradigmatic patterns, by reduplicating a few symbol clusters to encode symbols in a time-sensitive way.

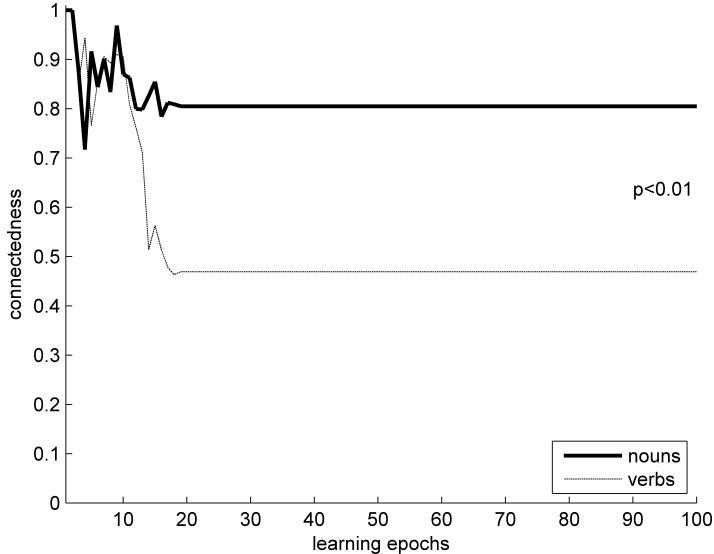


FIGURE 2. CONNECTEDNESS VALUES FOR TSOMs TRAINED ON NOUNS (THICK LINE) AND VERBS (THIN LINE) ACROSS LEARNING EPOCHS

Furthermore, we observe a statistically significant difference between singular and plural word forms in the converging epoch for recall accuracy, i.e. the learning epoch at which the map reaches its stable value in recalling a particular word class accurately (calculated as the learning epoch after which each word form is always recalled correctly). Figure 3 (next page) shows the different distribution of recall accuracy through training epochs (one hundred epochs), averaged over 5 map instances, for singular word forms (boxplot on the left) and plural word forms (boxplot on the right). Plural word forms show to be delayed in acquisition compared to their singular base forms, in line with psycholinguistic and developmental research evidence pointing to difficulties in acquisition and errors in overapplication and misapplication of the many plural marking patterns of German noun morphology.

In particular, despite an early onset of plural marking in German, empirical studies investigating spontaneous child speech (Szagun 2001; Laaha et al. 2006; among others) show differences in the acquisition rates for different plural suffixes, though error rates remain high during the whole pre-school age. Results show an undoubted advantage for the acquisition of the *-(e)n* plural suffix pattern, followed by *-s*, whereas the umlauting process

without suffix presents the longest period of instability and misapplication. Results on all other plural patterns seem to be less converging. The choice of plural marking depends on several aspects, such as gender as well as phonological and prosodic constraints (Wegener 1995; Wiese 1996). More interesting for our present concerns is the role of frequency and productivity.

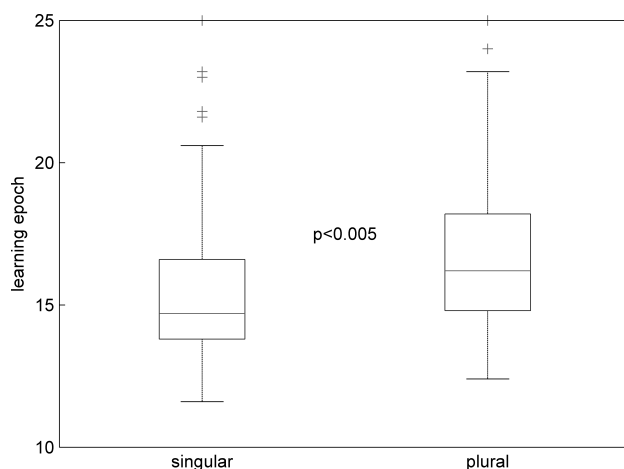


FIGURE 3. DISTRIBUTION OF THE CONVERGENCE EPOCH FOR RECALL ACCURACY ON 250 NOUN SINGULARS (LEFT BOXPLOT) AND 250 PLURALS (RIGHT BOXPLOT)

The  $-(e)n$  plural pattern is the most frequent one, namely about the 96% of German feminine nouns, and masculine nouns ending with  $-e$  (a phonological schwa), are marked by this plural pattern. Coherently, empirical findings show how this plural marker is used as the dominant suffix for overgeneralisation in spontaneous speech (Laaha et al. 2006; Niedeggen-Bartke 1999). On the other hand, other studies based on elicitation tasks show widespread overapplication of the  $-(e)n$  and  $-s$  patterns especially to plural nouns marked with umlaut without suffix (Niedeggen-Bartke 1999; MacWhinney 1978; among others).

In any case, the over-application of the  $-s$  pattern as a plural marker, together with more frequent plural suffixes, leads to the conclusion that frequency alone does not completely account for the choice of plural markers.

#### 4.2 Input factors: the effect of type/token frequency

In a usage-based perspective, frequency and consistency in morpho-syntactic structures are determining factors in language acquisition rates (Lieven & Tomasello 2008; Tomasello 2003). Input distributional properties, namely type and token frequency, play a crucial role in increasing learning accuracy (Bybee 2002, 2003).

High token-frequency word forms are memorised and accessed significantly earlier than low frequency forms; isolated representatives of inflectional classes (hapax or nearly hapax class members) are more prone/vulnerable to extension to other classes than members of densely populated classes. On the other hand, high type frequency and more regular classes are produced and generalised more easily.

To verify the impact of token frequency on accuracy scores and its effect on learning rates, in the second experiment we selected 300 of the 500 top frequency ranked noun word forms, and administered them to a 35×35 nodes map with their actual token frequencies. As reported in Figure 4, token frequency has a strong influence on early stages of language acquisition, in both recoding and recalling. In line with a coherent body of empirical evidence (Kauschke et al. 2011; Korecky-Kröll & Dressler 2009; among others), the first markers of plural to be acquired correspond to the most frequent input patterns.

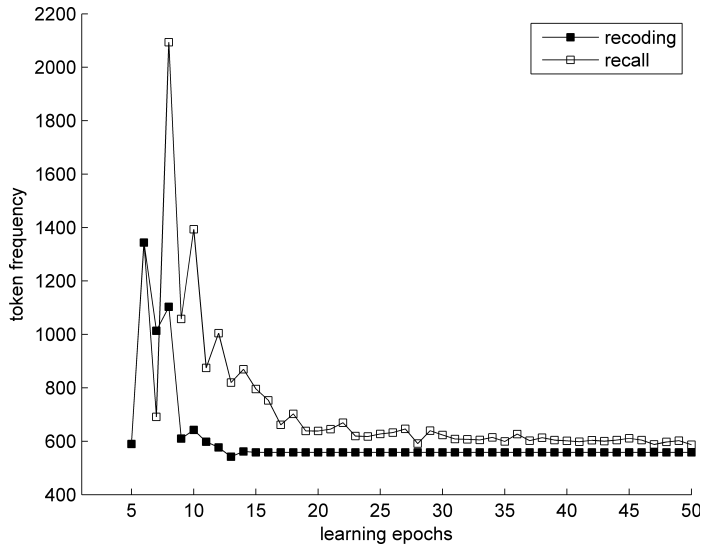


FIGURE 4. AVERAGE TOKEN FREQUENCY OF CORRECTLY RECODED (BLACK CIRCLES) AND RECALLED (WHITE CIRCLES) WORD FORMS (BOTH SINGULAR AND PLURAL) OVER THE FIRST FIFTY TRAINING EPOCHS

By monitoring in detail the *-en* and *-s* plural classes only, we report in Figure 5 the average token frequency of correctly recoded and recalled words plotted during learning epochs. It can be appreciated an earlier acquisition of the more frequent *-en* class compared to the less attested *-s* plural class. Token frequency has an impact both on correctly recoding and on accuracy in recalling: dark and white circles indicate respectively recoding and recalling mean token values of *-en* word forms (both singular and plural); dark and white triangles indicate values for *-s* class nouns.

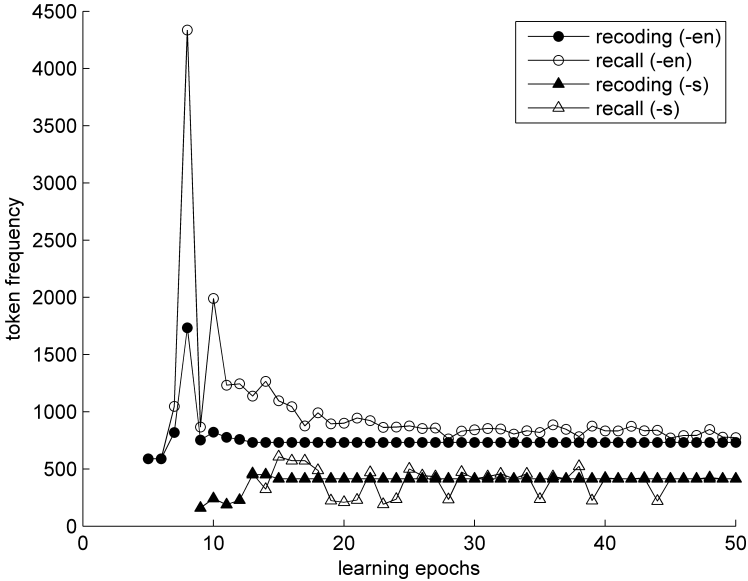


FIGURE 5. AVERAGE TOKEN FREQUENCY OF CORRECTLY RECODED AND RECALLED SINGULAR AND PLURAL FORMS OF *-EN* AND *-S* CLASSES

### 4.3 Language specific factors: predictability and transparency

In TSOMs (as well as in usage-based models), highly type-frequent inflection classes are expected to be most productive, productivity being however modulated by within-class token distribution and regularity of class members.

As already discussed (Section 2.1), evenly-distributed, low-frequency words showing a regular morphological pattern tend to share overlapping node chains, with the first node in a shared chain being densely inter-connected with other nodes through highly-entropic pre-synaptic connections. Since entropy increases with the number of chain-sharing words, we expect more regular and evenly-distributed plural classes to develop higher entropic pre-synaptic connections at morpheme boundaries. Finally, densely interconnected nodes define potentially productive morphological schemas, which are then used by novel words when they happen to share the same regular pattern.

To verify such a broad range of predictions, and their relation to the formal integrity of the base forms, we calculated the entropy of the distribution of weights on pre-synaptic connections at morpheme boundary, for each of the plural classes requiring a suffix. The plot (Figure 6) shows a significantly ( $p < 0.01$ ) higher entropy for the *-s* plural class. This means that, on average, more *-s* plurals tend to share overlapping nodes than any other plural classes. This is due to several factors: first, *-s* plurals tend to cluster in consistent families (e.g. phonologically consistent sub-groups, such as base forms



ending in vowels – as for example *AUTO*, *BÜRO*, *RADIO*, *FOTO* – or forms exhibiting untypical phonotactic patterns which make them be perceived as dissimilar from forms of all other noun classes – as for example *TAXI*, *SAISON*), which contributes to make them particularly reliable predictors of class membership. Furthermore, *-s* plurals are evenly distributed within their own inflectional class, a factor that minimises inter-node competition and favours sharing of memory resources. Finally, we believe that another contributing factor is perception of morphological structure, i.e. the extent to which *-s* is parsed and perceived by the map as an inflectional marker.

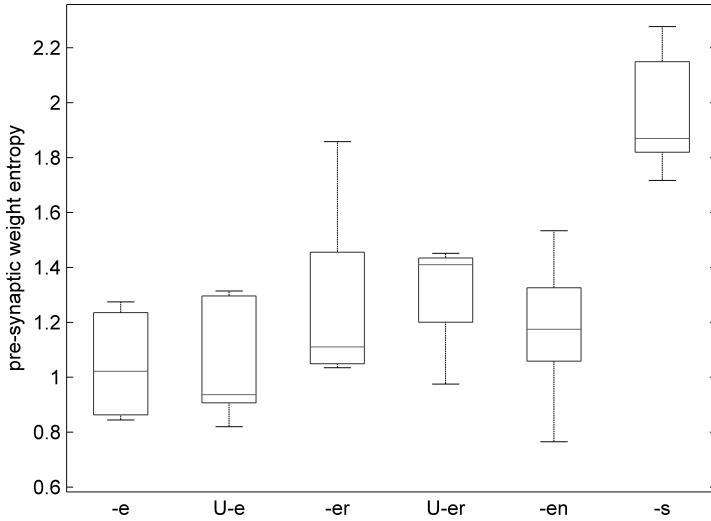


FIGURE 6. PER-CLASS DISTRIBUTION OF WEIGHTS ON PRE-SYNAPTIC CONNECTIONS AT MORPHEME BOUNDARY

To quantify perception of morphological structure, Figure 7 (next page) shows the distribution of maximum entropy values over pre-synaptic connections at different distances from the stem-ending boundary. It should be noted that only the *-s* class has many plurals which maximise entropy values at the morpheme boundary. For all other classes, the map is perceiving a morpheme boundary most often only after the first element of the plural suffix. In fact, in German nouns stem-final *-(e)n*, *-e*, *-er* are also found in a few singular forms (e.g. *WAGEN*, *BEIN*, *BRUDER*, *FINGER*, *NUMMER*, *BIER*, *ANFRAGE*, *IDEE*, *SEKUNDE*), and the same suffixes are markers of other morphological phenomena (e.g. verb forms, comparative adjective), thus reducing the iconicity of these segments as plural markers. On the other hand, the *-s* is the most unambiguous plural marker, because of the total rarity of singular nouns ending in *-s* (Wegener 1995).

To sum up, although relatively infrequent, *-s* plurals seem to pattern in fairly regular sub-classes, which suffer from no – or little – competition by members of other inflectional classes. As a result, the *s*-ending tends to be

perceived by the map as a point of discontinuity in the morphological structure of *-s* plurals, with weak, highly-entropic connections at morpheme boundaries, more than in any of the other plural classes.

Finally, the *s*-ending does not affect the formal integrity of the corresponding singular base, thus requiring no acquisitional overhead.

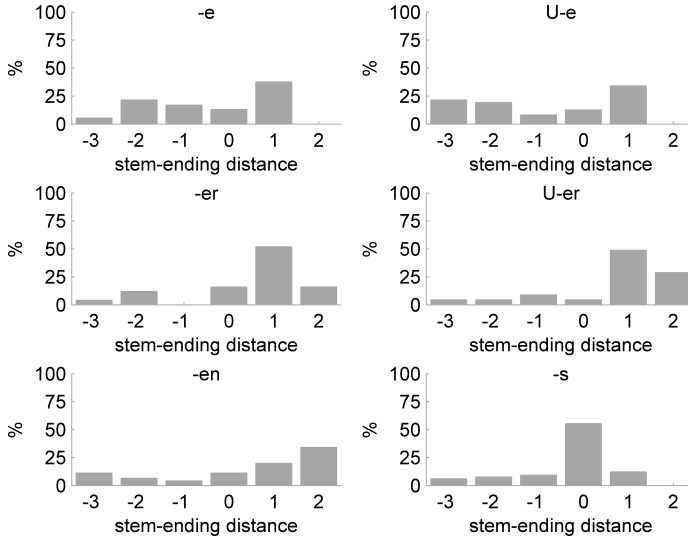


FIGURE 7. PER-CLASS DISTRIBUTION OF MAXIMUM ENTROPY VALUES OF PRE-SYNAPTIC CONNECTIONS AT DIFFERENT DISTANCES FROM THE STEM-ENDING BOUNDARY

## 5. DISCUSSION AND CONCLUDING REMARKS

Acquisition rate is subject to input factors, but with some qualifications. It is commonly assumed that language-specific factors have an influence on acquisition and generalisation. In addition to frequency, other concurrent, language specific factors are determinant in the acquisition of plural markers, namely the formal consistency and reliability of inflectional patterns.

In the frame of empirical studies on German plural markers acquisition, some aspects have been identified as having an influence on the acquisition rate and selection of possible plural patterns. The evidence shows a correlated impact of suffix predictability and stem transparency on acquisition of plurals (Laaha & Dressler 2012), as well as of the unambiguity of plurality marking, referred to as *ICONICITY* of plural markers (Zaretsky et al. 2011; Bittner & Köpcke 2001). The plural patterns without umlaut are more iconic than those with umlaut, and the schema umlaut with suffix zero is not iconic at all. The more iconic and morphological transparent a plural marker is, the earlier it is acquired and overgeneralised (Korecky-Kröll & Dressler 2009).

Although 1-layer TSOMs cannot account for all factors governing the

productivity of inflectional classes, we showed how they are instrumental in highlighting a few interesting formal properties of the *-s* inflectional class: far from exhibiting a default behaviour, the *s*-pattern appears to apply to a larger stem space than other more frequent pluralisation patterns in German noun inflection.

The ontogenetic trajectory of the acquisitional process we outlined here is the result of an ADAPTIVE MULTI-FACTORIAL PROCESS, where token frequency (as highlighted in our second experiment, Section 4.2) has a strong impact on early acquisitional stages, accounting for the piecemeal acquisition of item-based data (in line with UB models); whereas type frequency strongly correlates with the emergence of recurrent sequences which represent a key element in determining productivity and generalisation of abstract patterns of sequential arrangement (Bybee 2002; Tomasello 2006). In addition to, and not in contradiction with this account, besides distributional properties, other concurrent factors, relying on salience and iconicity of inflectional markers, account for the dynamic emergence of morphology acquisition and processing. Our results show (Section 4.3) how suffix predictability and stem transparency contribute to the perception of morphological structure.

As observed by Hay & Baayen (2002), not all morphologically complex words contribute to productivity equally. Only words which are perceived as morphologically complex by being parsed, contribute to high levels of activation of their sublexical constituents and to activation spreading to other words sharing the same constituents in the mental lexicon. This is strongly supported by our analysis of the behaviour of TSOMs, where word representations are dynamically recoded time-series, whose perception can vary depending on the context and on recurrent patterns of underlying morphological structure. In our framework, morphological structure and generalisation are analysed as a by-product of flexible memory self-organisation strategies for word recoding.

## REFERENCES

- Alegre, M. & P. Gordon (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40. 41-61.
- Anderson, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.
- Baayen, H., R. Piepenbrock & L. Gulikers (1995). *The CELEX lexical database* (CD-ROM). Philadelphia: Linguistic Data Consortium.
- Baayen, R. H., R. Lieber & R. Schreuder. (1997). The morphological complexity of simplex nouns. *Linguistics* 35. 861-877.
- Baayen, R. H., H. L. Wurm & J. Aycok (2007). Lexical dynamics for low-frequency complex words. A regression study across tasks and modalities. *The Mental Lexicon* 2(3). 419-463.

- Bartke, S., F. Rösler, J. Streb & R. Wiese (2005). An ERP-study of German “irregular” morphology. *Journal of Neurolinguistics* 18. 29-55.
- Bates, E. & B. MacWhinney (1987). Competition, variation, and language learning. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, 157-193. Mahwah, NJ: Erlbaum.
- Bittner, D. & K.-M. Köpcke (2001). Acquisition of the German plural markings: A case study in natural and cognitive morphology. In C. Schaner-Wolles, J. Renison & F. Neubarth (eds.), *Naturally! Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of his 60th birthday*, 47-58. Turin: Rosenberg & Sellier.
- Blevins, J. P. (2006). Word-based morphology. *Journal of Linguistics* 42. 531-573.
- Bybee, J. (1985). *Morphology. A study of the relation between meaning and form*. Amsterdam & Philadelphia: John Benjamins.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425-455.
- Bybee, J. (2002). Sequentiality as the basis of constituent structure. In T. Givón & B. Malle (eds.), *The evolution of language out of pre-language*, 107-132. Amsterdam & Philadelphia: John Benjamins.
- Bybee, J. (2003). Cognitive processes in grammaticalization. In M. Tomasello (ed.), *The new psychology of language*, Vol. II. New Jersey: Erlbaum.
- Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition*, 216-236. New York: Routledge.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, J. & P. Hopper (eds.) (2001). *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: John Benjamins.
- Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences* 22. 991-1060.
- Croft, W. (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.
- Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition* 24. 143-188.
- Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1-24.
- Ferro, M., G. Pezzulo & V. Pirrelli (2010). Morphology, memory and the mental lexicon. *Lingue e Linguaggio* 9(2). 199-238.
- Ferro, M., C. Marzi & V. Pirrelli (2011). A self-organizing model of word storage and processing: Implications for morphology learning. *Lingue e Linguaggio* 10(2). 209-226.
- Hahn, U. & R. C. Nakisa (2000). German inflection: Single route or dual route? *Cognitive Psychology* 41. 313-360.
- Haspelmath, M. (2002). *Understanding morphology*. London: Arnold.
- Hay, J. B. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics* 39. 1041-1070.
- Hay, J. B. & R. H. Baayen (2002). Parsing and productivity. In G. E. Booij & J. van Marle (eds.), *Yearbook of Morphology 2001*, 203-235. Dordrecht: Kluwer.

- Jackendoff, R. (2002). *Foundations of language. Brain, meaning, grammar, evolution*. New York: Oxford University Press.
- Kauschke C., A. Kurth & U. Domahs (2011). Acquisition of German noun plurals in typically developing children and children with specific language impairment. *Child Development Research*. Article ID 718925.
- Kohonen, T. (2001). *Self-organizing maps*. Berlin: Springer.
- Korecky-Kröll, K. & U. W. Dressler (2009). The acquisition of number and case in Austrian German nouns. In U. Stephany & M. D. Voeikova (eds.), *Development of nominal inflection in first language acquisition: A cross-linguistic perspective*, 265-302. Berlin & New York: Walter de Gruyter.
- Koutnik, J. (2007). Inductive modelling of temporal sequences by means of self-organization. In *Proceeding of International Workshop on Inductive Modelling (IWIM 2007)*, 269-277. Ljubljana: CTU in Prague.
- Laaha, S., D. Ravid, K. Korecky-Kroll, G. Laaha & W. U. Dressler (2006). Early noun plurals in German: Regularity, productivity or default? *Journal of Child Language* 33(2). 271-302.
- Laaha, S. & W. U. Dressler. (2012). Suffix predictability and stem transparency in the acquisition of German noun plurals. In F. Kiefer, M. Ladányi & P. Siptár (eds.), *Current issues in morphological theory: (Ir)regularity, analogy and frequency*, 217-235. Amsterdam & Philadelphia: John Benjamins.
- Lieven, E. & M. Tomasello (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition*, 216-236. New York: Routledge.
- MacWhinney, B. (1978). The acquisition of morphophonology. In *Monographs of the Society for Research in Child Development* 43. 1-123.
- MacWhinney, B. (1987). The competition model. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, 249-308. Mahwah, NJ: Erlbaum.
- Marcus, G. F., U. Brinkmann, H. Clahsen, R. Wiese & S. Pinker (1995). German inflection: The exception that proves the rule. *Cognitive Psychology* 29. 189-256.
- Marzi, C., M. Ferro & V. Pirrelli (2012a). Prediction and generalisation in word processing and storage. In *On-line Proceedings of the Eighth Mediterranean Morphology Meeting (MMM8)*, 113-130.
- Marzi, C., M. Ferro & V. Pirrelli (2012b). Word alignment and paradigm induction. *Lingue e Linguaggio* 11(2). 251-274.
- Marzi, C., M. Ferro, C. Caudai & V. Pirrelli (2012c). Evaluating Hebbian self-organizing memories for lexical representation and access. In *Proceedings of 8<sup>th</sup> International Conference on Language Resources and Evaluation (ELRA – LREC 2012)*, 886-893.
- Moscoso del Prado Martín, F. (2007). Co-occurrence and the effect of inflectional paradigms. *Lingue e Linguaggio* 6(2). 247-262.
- Moscoso del Prado Martín, F., A. Kostić & H. Baayen (2004a). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition* 94. 1-18.
- Moscoso del Prado Martín, F., R. Bertram, T. Häikiö, R. Schreuder & H. Baayen (2004b). Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental*

- Psychology: Learning, Memory and Cognition* 30(6). 1271-1278.
- Niedeggen-Bartke, S. (1999). Flexion und wortbildung im spracherwerb. In J. Meibauer & M. Rothweiler (eds.), *Das Lexikon im Spracherwerb*, 208-228. Tübingen: Francke.
- Pinker, S. & A. Prince (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 29. 195-247.
- Pinker, S. & M. T. Ullman (2002). The past and future of the past tense. *Trends in Cognitive Science* 6. 456-463.
- Pirrelli, V., M. Ferro & B. Calderone (2011). Learning paradigms in time and space. Computational evidence from Romance languages. In M. Goldbach, M. O. Hinzelin, M. Maiden & J. C. Smith (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 135-157. Oxford: Oxford University Press.
- Plunkett, K. & R. C. Nakisa (1997). A connectionist model of the Arabic plural system. *Language and Cognitive Processes* 12(5/6). 807-836.
- Prasada, S. & S. Pinker (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* 8. 1-56.
- Rumelhart, D. & J. McClelland (1986). On learning the past tense of English verbs. In D. E. Rumelhart & J. McClelland (eds.), *Parallel distributed processing: Explanations in the microstructure of cognition*. Cambridge, MA: The MIT Press.
- Sonnenstuhl, I. & A. Huth (2002). Processing and representation of German *-n* plurals: A dual mechanism approach. *Brain and Language* 81. 276-290.
- Szagan, G. (2001). Learning different regularities: The acquisition of noun plurals by German-speaking children. *First Language* 21(62). 109-141.
- Stump, G. T. (2001). *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press.
- Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition* 7. 263-272.
- Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2006). Acquiring linguistic constructions. In D. Kuhn & R. Siegler (eds.), *Handbook of child psychology*. New York: Wiley.
- Wegener, H. (1995). *Die Nominalflexion des Deutschen: Verstanden als Lerngegenstand*. Tuebingen: Niemeyer.
- Wiese, R. (1996). *The phonology of German*. Oxford: Oxford University Press.
- Zaretsky, Y., B. P. Lange, H. A. Euler & K. Neumann (2011). Pizzas, Pizzen, Pizze: Frequency, iconicity, cue validity, and productivity in the plural acquisition of German pre-schoolers. *Acta Linguistica* 5(2). 22-35.

*Claudia Marzi*

Istituto di Linguistica Computazionale  
 "A. Zampolli"  
 Consiglio Nazionale delle Ricerche  
 Via G. Moruzzi 1, 56124 Pisa  
 Italy  
 e-mail: claudia.marzi@ilc.cnr.it

*Marcello Ferro*

Istituto di Linguistica Computazionale  
 "A. Zampolli"  
 Consiglio Nazionale delle Ricerche  
 Via G. Moruzzi 1, 56124 Pisa  
 Italy  
 e-mail: marcello.ferro@ilc.cnr.it