Claudia Marzi, James P. Blevins, Geert Booij and Vito Pirrelli

Inflection at the morphology-syntax interface

Abstract: What is inflection? Is it part of language morphology, syntax or both? What are the basic units of inflection and how do speakers acquire and process them? How do they vary across languages? Are some inflection systems somewhat more complex than others, and does inflectional complexity affect the way speakers process words? This chapter addresses these and other related issues from an interdisciplinary perspective. Our main goal is to map out the place of inflection in our current understanding of the grammar architecture. In doing that, we will embark on an interdisciplinary tour, which will touch upon theoretical, psychological, typological, historical and computational issues in morphology, with a view to looking for points of methodological and substantial convergence from a rather heterogeneous array of scientific approaches and theoretical perspectives. The main upshot is that we can learn more from this than just an additive medley of domain-specific results. In the end, a cross-domain survey can help us look at traditional issues in a surprisingly novel light.

Keywords: Inflection, paradigmatic relations, word processing, word learning, inflectional complexity, family size, entropy

1 The problem of inflection

Inflection is the morphological marking of morphosyntactic and morphosemantic information like case, number, person, tense and aspect (among others) on words. For instance, a word may be specified as singular for the grammatical category of number, i.e. it has a certain value for the feature ‘number’. The feature ‘number’ has two values in English: singular and plural. The choice of specific values for such features may depend on syntactic context or semantic context.

Claudia Marzi, Vito Pirrelli, Italian National Research Council (CNR), Institute for Computational Linguistics, Pisa, Italy

James P. Blevins, George Mason University, College of Humanities and Social Sciences, Fairfax, USA

Geert Booij, Leiden University, Leiden University Centre of Linguistics, Leiden, Netherlands

Morphosyntactic features are inflectional features that play a role in syntax. That is, they play an essential role in the interface between morphology and syntax. For instance, the syntax of languages often requires that words in specific syntactic contexts agree with respect to the value for certain features of other, syntactically related words. An example is subject-verb agreement in English: the finite form of a verb has to agree in the values for person and number with those of the subject. Another well-known type of agreement is gender agreement: in many languages determiners and modifying adjectives have to agree in gender with their head noun.

Besides agreement, there is a second type of syntactically driven feature value selection, traditionally referred to as government: a word or syntactic construction is said to govern the choice of a feature value for another word. For instance, in many languages nouns have to be marked for case, depending on the syntactic or semantic role of that noun (subject, object, agent, etc.). The grammatical or semantic role of an NP then governs the case marking of its head noun.

Morphosemantic features are features that are not required by a syntactic context, and their choice is primarily motivated semantically. For example, all finite forms of English have a tense property such as present or past. Their choice is not governed by syntactic context, but by what content the speaker wants to convey. Yet, the choice is obligatory, as a specific tense property has to be chosen. In this respect, inflection differs from derivation, which is not obligatory. However, context may play a role in the choice of morphosemantic features as well. For instance, in a sentence such as *Yesterday I went to the movies* the past tense form *went* is normally required because of the presence of the adverb *yesterday*.

Inflection may be divided into two subtypes: *inherent inflection* and *contextual inflection* (Booij 1993, 1996; Kibort 2010). Inherent inflection is primarily determined by what the speaker wants to express, and is therefore a matter of choice. The speaker determines, for example, the choice between present tense and past tense of verb forms, and the choice of number (singular or plural) for nouns. Contextual inflection is the kind of inflection that is required by syntactic context. This is the case for the choice of person and number values for finite verbs in English, which is a matter of agreement. Hence, in English the feature ‘number’ is inherent for nouns, but contextual for verbs. Case marking on nouns may function as contextual inflection in a language like German, where subject nouns and object nouns have to be marked as having nominative and accusative case respectively. That is, this instance of case marking is required by syntax. These are the so-called structural cases, and stand in contrast with semantic case marking, which is a case of inherent inflection. For instance, in Latin we can express the instrumental use of a knife by means of marking the noun with ablative case: *cultr-o* ‘with a knife’. This is a semantically governed

case, and hence a matter of inherent inflection. Adjectives in German agree in case marking with their head nouns, and, therefore, case marking is contextual. A survey of the different types of inflectional features is given in Kibort (2010).

An example that illustrates what can be expressed by inflection is the following sentence of the language Maale, a North Omotic language spoken in South Ethiopia (Amha 2001: 72):

- (1) bayí-ské-nn-ó   tá   zag-é-ne
       cow-INDF-F-ABS 1SG.NOM see-PRF-AFF.DECL
       ‘I saw a cow (which I did not know before)’

The word for *cow* has morphological markers for indefiniteness, feminine gender, and absolutive case, and the verb is marked for aspect (Perfect) and for the sentence being affirmative (AFF) and declarative (DECL) in meaning. The ending *-ne* is the cumulative exponence of the two sentence modalities Affirmative and Declarative. The pronoun for ‘I’ is the nominative form, but there is no separate case marker. This example illustrates some of the formal complications in the expression of inflectional categories, and in particular that there may be no one-to-one mapping of form and meaning in inflection. Inflection may thus considerably increase the formal complexity (discussed in Section 8) of a language system.

A third type of traditional inflectional features are purely morphological features such as inflectional classes for nouns and verbs. In Latin, case marking on nouns is performed in five different ways, and hence it is conventional to distinguish five different inflectional classes (declensions) for nouns. Individual nouns are then marked for the inflectional class they belong to by means of a feature. These features are purely morphological because they tend to have no role in syntax or semantics. Similarly, Latin is associated with a number of inflectional classes for verbs, the conjugations. The patterns described in terms of inflectional classes add substantially to the complexity of a language, and raise the question how children acquire such morphological systems.

In many languages the gender of nouns is marked on related words. For instance, in Dutch nouns have either common or neuter gender, and this manifests itself in agreement phenomena: determiner and adjective have to agree in gender and number with the head noun. However, the nouns themselves do not carry a morphological marker for gender. Thus, one can only discover the gender of Dutch nouns indirectly, by looking at agreement data. This is another challenge for the language learner.

Two interacting but logically distinct issues lie at the core of any linguistic or psycholinguistic account of inflection:

- A. the issue of what syntactic contexts require morphosyntactic and/or morpho-semantic word marking and for what lexical/grammatical units;
- B. the issue of how morphosyntactic and morpho-semantic information is overtly realized on lexical/grammatical units.

In this chapter we mainly focus on the second issue, the ways in which morpho-syntactic and morphosemantic information is morphologically marked. It must be observed that there is no one-to-one relationship between inflectional features and units of form ('morphs') in inflected words, as we will see in Section 2: one morph may express more than one inflectional property (cumulative exponence), and one inflectional property may be expressed by more than one feature (extended exponence). Moreover, the same inflectional property may be expressed in a number of different ways. The patterns of interdependent choices are expressed by inflectional classes. A simple example from English is that the past tense forms of verbs may be formed either by means of suffixation of the stem with *-ed*, or by means of various types of apophony (Ablaut), i.e. vowel change in the stem, usually with consequences for the form of participles. It does not make any difference for the role of the feature value 'Past' in syntax and semantics by which formal means it is expressed. This issue is broached in more detail in Section 2.

Given the lack of a one-to-one mapping of form and meaning in the domain of inflection, it is useful to introduce paradigms, systematically structured sets of inflectional forms of words, in order to make the right generalizations and the proper computations (discussed in Section 5). The nature and structure of inflectional paradigms are discussed in Section 3.

The inflectional paradigms of words may also contain word combinations. For instance, in Germanic and Romance languages various inflectional forms can be treated as consisting of an auxiliary and a non-finite form of a verb. This is referred to as periphrasis (Section 4).

After this brief sketch of the nature of inflectional systems, we will discuss in more detail the way inflection is acquired, how machines can learn it, how it can be modeled computationally and how morphological complexity can be computed (Sections 5–8). Section 9 will summarize our findings.

2 Inflectional syntagmatics

The inflectional features of a morphological system determine observable patterns of variation in shape and distribution. In turn, observable patterns of variation cue features. Yet the relations between features and patterns of variation

are often intricate, typically involving complex interactions between syntagmatic arrangements and paradigmatic classes.

2.1 Morphemes and inflection

Post-Bloomfieldian models seek to establish a tight connection between inflectional features and morphotactic units by bundling features and forms into inflectional ‘morphemes’. In the immediate constituent analyses developed by Bloomfield’s successors, inflectional formatives were included among the terminal elements of a syntactic representation, along with bound stems and free forms. The idea of treating inflectional sub-word units as syntactic elements was taken over by generative accounts, leading to the notion of ‘functional categories’ and to a general conception of morphology as the ‘syntax of words’.

These inflectional sub-word ‘units’ have long presented some of the most stubbornly recalcitrant challenges for morphemic analysis. Initial attempts to align individual inflectional features with morphotactic units created analytical conundrums in languages as inflectionally impoverished as English. Harris (1942: 113) and Hockett (1947: 240) struggled with the task of segmenting English *children* into morphemic units, due to uncertainty about the synchronic status of the historical strong (-r) and weak (-en) plural markers. Cognate patterns in other West Germanic languages raise similar problems. For example, some nouns in Modern German distinguish singulars and plurals by an ending, as illustrated by *Tag~Tage* ‘day(s)’. Other nouns mark the contrast by a medial vowel alternation as in *Garten~Gärten* ‘garden(s)’. In other nouns, the contrast is marked both by an ending and vowel alternation, as in *Fuchs~Füchse* ‘fox(es)’. In yet other nouns, such as *Kabel* ‘cable(s)’ there is no variation between the singular and plural. A biunique correspondence cannot be established in a uniform manner between the feature ‘plural’ and a ‘unit of form’ in these cases without assigning more abstract analyses to the surface forms.

Various technical strategies have been explored for assigning morphemic analyses to these and other seemingly non-biunique patterns of inflectional marking. One type of proposal generalizes the notion of ‘form’. Among the initial generalizations were different varieties of ‘special morphs’, such as the ‘process morphs’ that bundled pairs of alternating vowels into ‘units’. Modern descendants include morphophonemic ‘readjustment rules’ (Halle and Marantz 1993), which intervene between morphemic analyses and surface forms. An alternative strategy involves reclassifying patterns of surface variation, so that some markers can be discounted in determining morphemic biuniqueness (Noyer 1992).

Contemporary interest in exploring these types of technical refinements tends to be concentrated in communities with an overarching commitment to a syntactico-centric conception of morphology. From a morphological perspective, the core problems of morphemic analysis derive from an excessively narrow view of morphological structure and, as such, do not seem amenable to purely technical solutions. Instead, as argued in Matthews (1972, 1991), biunique relations are best understood as limiting cases of generally many-to-many relations between inflectional features and units of form. The extended discussion of Latin conjugational patterns in Matthews (1972) traces the problems of segmentation and interpretation created by coercing a morphemic analysis onto languages that do not conform to an agglutinative ideal. Many-to-many relations between features and forms are so endemic to Latin and Ancient Greek that they are robustly exhibited by regular and even exemplary items. To illustrate this point, Matthews (1991: 174) considers the Ancient Greek form *elelykete* (e-le-ly-k-e-te) ‘you had unfastened’, which, as he notes, does not show “any crucial irregularity” and “is in fact the first that generations of schoolchildren used to commit to memory”:

But categories and formatives are in nothing like a one-to-one relation. That the word is Perfective is in part identified by the reduplication *le-* but also by the suffix *-k-*. At the same time, *-k-* is one of the formatives that help to identify the word as Active; another is *-te* which, however, also marks it as ‘2nd Plural’.

(Matthews 1991: 173)

The deeper problem, as Matthews emphasizes, is not just that ‘flectional’ languages like Latin or Greek appear to exhibit morphemic indeterminacy, but that the indeterminacy is the artifact of a method. By foisting an agglutinative analysis onto flectional languages, a morphemic approach creates problems for which it can provide no principled solution. The attempt to address these problems through technical refinements of a morphemic model seems futile; at least some languages falsify the assumptions of the model:

One motive for the post-Bloomfieldian model consisted, that is to say, in a genuinely factual assertion about language: namely, that there is some sort of matching between minimal ‘sames’ of ‘form’ (morphs) and ‘meaning’ (morphemes). Qua factual assertion this has subsequently proved false: for certain languages, such as Latin, the correspondence which was envisaged apparently does not exist . . . One is bound to suspect, in the light of such a conclusion, that the model is in some sense wrong.

(Matthews 1972: 124)

Subsequent studies have provided further confirmation that inflectional features and units of form are, in the general case, related by many-to-many ‘exponent’ relations. One strand of this research has even explored relations that are more ‘exuberantly’ many-to-many than the patterns exhibited by classical languages (Harris 2009; Caballero and Harris 2010). A pair of related conclusions can be

drawn from this work. The first is that, when applied to all but the most uniformly agglutinative structures, morphemic models create problems of analysis while also obscuring the organization and function of form variation. The second is that morphemes cannot provide the basis for inflectional description and that even descriptive conventions like ‘morpheme glosses’ harbor untenable idealizations.

2.2 Varieties of exponence

Although exponence relations do not share the descriptive shortcomings of morphemes, they exhibit their own characteristic limitations. What unites the diverse exponence relations investigated in the realizational literature is a fundamentally negative property: the relations all involve non-biunique feature-form associations. The realizational tradition offers no positive characterization of these relations, and contains almost no discussion of what functions, if any, might be associated with different patterns of exponence.

A model that recognizes many-to-many feature-form relations sacrifices the attractively simple compositional semiotics of a morphemic model, in which complex forms and complex meanings are built up in parallel from atomic form-meaning pairs. As expressed by the ‘Separation Hypothesis’ (Beard 1995), realizational accounts treat feature bundles as ‘minimum meaningful units’ and assign no meaning or function to variation in the ‘spell-out’ of bundles. In effect, realizational models move from one extreme to another, replacing individually meaningful morphemes with collectively meaningless exponents.

Both of these extremes reflect a set of limiting assumptions about the nature of inflectional functions and meanings. Three assumptions are of particular importance. The first is that meanings are exclusively ‘extramorphological’ and do not encode information about the shape or distribution of related forms, or other properties of the morphological system. The second is that discrete meanings are associated statically with forms, not determined dynamically within a network of contrasts. The third is that analyses and interpretations are taken to be assignable to forms in isolation from the systems in which they function.

By incorporating these assumptions, realizational approaches close off any inquiry into the meaning or function of different patterns of inflectional exponence. The adoption of other, equally conservative, assumptions imposes similarly severe constraints. Whereas the role of paradigmatic structure has been a matter of dispute for most of the modern period, the relevance of morphotactic structure has remained almost unquestioned. However, in even the best-studied languages, there have been no systematic attempts to provide cognitive motivation for the morphotactic structures assigned in standard descriptions. It is

sometimes believed that a shift from orthographic to phonemic representation places analyses on a firmer foundation. But it is well-known that inflectional contrasts may be cued by sub-phonemic contrasts (Baayen et al. 2003; Kemps et al. 2005). Moreover, the crude procedures of distributional analysis used to determine morphotactic structure have no provision for distinguishing sequences that function as units in the synchronic system from those that are diachronic relics of the processes of morphologization that produced the system.

2.3 Synchronic vs diachronic structure

The confound between synchronically active structure and diachronic residue derives ultimately from the general conflation of synchronic and diachronic dimensions in the post-Bloomfieldian tradition. To a large extent, this tradition is guided by the goal of exploring the descriptive potential of recasting diachronic analyses in synchronic terms. Underlying representations are transparent synchronic proxies for the ‘least common ancestors’ of a set of surface forms. Due to the regularity of sound changes, these ancestors will tend to be definable in terms of a minimum edit distance from their descendants. The serial structure of derivations likewise mirrors the temporal order of the sound changes that occur in the history of a language. The agglutinative bias of a morphemic model (termed ‘The Great Agglutinative Fraud’ by Hockett (1987: 83)) also reflects an essentially historical perspective. A perfectly agglutinative system is one in which the waves of grammaticalization that build up complex forms preserve the discrete meaning and morphotactic separability of the morphologized parts, while maintaining a link between their meaning and form.

However, the shift from a diachronic to a synchronic perspective is as disruptive to the interpretation of a post-Bloomfieldian model as the move from biuniqueness to many-to-many exponence relations is for the semiotics of a realizational approach. Morphotactic structure is often a useful guide to the historical processes that applied to produce the inflected forms of a language. Yet knowledge of the historical origins of forms can distort analyses of their current status and function. This problem arises in an acute form in the ‘templatic’ analyses assigned to languages of the Algonquian and Athapaskan families. One tradition of analysis, illustrated in the Navajo grammar of Young and Morgan (1987), associates verbs with morphotactic ‘templates’ consisting of sequences of ‘slots’, each containing a substitution class of formatives. An intricate overlay of distributional and interpretative dependencies holds between the ‘choices’ at different points in the template. Not all slots need be ‘filled’ in with a surface form and, typically, many are empty. As a result, there is a vast

contrast between the complexity of the ‘underlying’ templatic structure and the morphotactics of surface forms. In effect, the dimensions of variation in the templatic descriptions provide a record of the history of the derivation of the modern forms. The apparent ‘choices’ are no longer independent in the modern languages but have been encapsulated in larger recurrent sequences, as represented in the type of bipartite structure proposed by McDonough (2000).

In sum, despite the appeal of a uniform approach to syntactic and inflectional patterns, the models developed within the broad post-Bloomfieldian tradition incorporate biases and assumptions that have mostly impeded attempts to understand the structure and organization of inflectional systems. The analytical assumptions incorporated in approaches that were developed expressly to model inflectional patterns offer a usefully different perspective.

3 Paradigmatic distribution and interpretation

As their name suggests, classical ‘Word and Paradigm’ (WP) models shift the fundamental part-whole relation in an inflectional system onto the relation between individual words and inflectional paradigms. This shift does not deny the significance of sub-word variation but instead interprets variation in a larger paradigmatic context. The context is defined by the interaction of two main relational dimensions. The first dimension is **discriminative**, implicit in the etymological origin of the term ‘inflection’, which, as Matthews (1991:190) notes, derives “from a Latin Verb whose basic meaning was ‘to bend’ or ‘to modify’”. Patterns of form variation serve to distinguish larger ‘free forms’ with an independent distribution that can communicate meaning or function within a system. In contrast to morphemic models, variants are not individually meaningful; in contrast to realizational models, they are not collectively meaningless. The variants that discriminate a form determine its place in a similarity space defined by the inflectional contrasts exhibited by a language.

The second dimension is **implicational** or **predictive**, expressed by “the . . . general insight . . . that one inflection tends to predict another” (Matthews 1991: 197). Patterns of variation within an inflectional system tend to be interdependent in ways that allow speakers to predict novel forms on the basis of forms that they have encountered. The predictive patterns that hold between forms determine the place of a form within implicational networks.

Discriminative and implicational relations are particularly relevant to the organization of inflectional systems, which exhibit highly uniform patterns of contrast and predictability. The inflected variants of an open-class item typically

define a closed, uniform feature space and a largely transparent semantic space. For an item of a given word class or inflection class, it is possible to specify the features that are distinctive for that item. This degree of uniformity and item-independence is what permits the strict separation of features and form in a realizational model. From even partial exposure to an inflectional system, speakers can arrive at reliable expectations about the number of forms of an item, assign interpretations to these forms, and even predict the shape of not yet encountered and potential variants.

The discriminative and implicational organization of inflectional systems is saliently reflected in the ways that these systems are described. The notion of a ‘morphological gap’ tends to be applied to cases in which predictable inflected forms are either missing or occur less frequently than expected (compared with the frequency of corresponding forms of other items of the same class). Unattested derivational formations are more rarely described as ‘gaps’, since a derivational system is not conceptualized as a closed, fully populated, space of forms. The notions ‘suppletion’, and even ‘syncretism’, also apply almost exclusively to inflected forms. Suppletion reflects unmet expectations about predictability and syncretism violates assumptions of discriminability. In the derivational domain, both patterns are usually treated as cases of ambiguity. Conversely, whereas derivational formations are often described as ‘established’, this term is rarely applied to individual inflected forms.

3.1 Words

The role of words and paradigms is a distinctive characteristic of classical WP descriptions of inflectional patterns. The use of word forms to exhibit variation accords with the view that “[t]he word is a more stable and solid focus of grammatical relations than the component morpheme by itself” (Robins 1959: 128). In support of this claim, paradigmatic approaches present individual case studies but no systematic attempt to explain **why** words provide a useful basis for inflectional description. Instead, word-based analyses of inflectional systems exploit the generally positive correlation between unit size and grammatical determinacy. In part, this correlation derives from the monotonic nature of determinacy. A fully determinate form may be composed of individually indeterminate parts. To take a simple example, the indeterminacy associated with an element such as English *-er*, which may mark comparative forms of adjectives or agentive nominals, is resolved in the word form *spammer*. In contrast, a set of fully determinate parts cannot be arranged into an indeterminate whole without contradicting the original assumption that they are determinate.

The modern development of a quantitative and cognitively-grounded discriminative approach offers a further perspective on this issue. As ‘free forms’ in the sense of Bloomfield (1933), words are the smallest units with an independently storable range of distributions and uses. As such, words are the smallest units with the kinds of individual ‘conditioning histories’ that are approximated by semantic vectors in models of distributional semantics such as Marelli and Baroni (2015). Correlating with these distributional and functional properties are measures of the uncertainty at word boundaries and the ‘informativeness’ of these boundaries, as operationalized by effects on compressability (Gertzen et al. 2016).

The interpretive determinacy of words also underlies the ‘morphological information’ that they express, i.e. information about the shape and distribution of other forms in a system. A form in isolation is of limited diagnostic value, reflecting the fact that such ‘pure’ forms are ecologically invalid abstractions. Speakers encounter forms in syntagmatic and paradigmatic contexts that guide their interpretation and support reliable inferences about related forms. The Paradigm Structure Constraints of Wurzel (1970) represent an early attempt to model this structure in terms of logical implication. Subsequent approaches, exploiting insights from the morphological processing tradition represented by Kostić et al. (2003) and Moscoso del Prado Martín et al. (2004), develop more robust information-theoretic measures of implicational structure. This approach has offered a useful perspective on questions concerning the structure and ‘complexity’ of patterns and classes, the nature of defectiveness and other properties of inflectional systems. Overall, the approach lends support to the traditional view that the inflectional component is not an unstructured inventory but, rather, a system, in which patterns of interdependency facilitate predictions about the whole system from a subset of its forms. Patterns of predictability also contribute to an understanding of the learnability of complex inflectional systems, and help to clarify the degree of variation in the complexity of descriptions of inflectional systems. In at least some cases, this variation reflects the intrinsic difficulty of describing systems as assemblages of independent items that are composed of recurrent parts. Although this may provide a reasonable basis for enumerating the patterns attested in a language by means of a written grammar, it is not a plausible model of acquisition or use. Much of the extreme cross-linguistic variation in grammars and typological accounts appears to be an artefact of descriptive tasks that never arise for speakers and hence are not subject to pressures that ensure learnability.

3.2 Paradigms

In descriptions of inflection class morphology, the variation exhibited by a class system is typically illustrated by full paradigms of ‘exemplary’ items, together with diagnostic ‘principal parts’ for non-exemplary items. There has been a tendency for this descriptive practice to be overinterpreted, by advocates as well as by opponents of paradigmatic approaches.

Reflecting the pedagogical origins of the classical tradition, exemplary paradigm and principal part descriptions are designed to provide a description of inflectional patterns that is maximally transparent for learners. There is neither cognitive nor empirical motivation for assuming that this pedagogical organization mirrors the format in which speakers represent knowledge about inflectional patterns. This observation can be traced at least to Hockett (1967), in speculations about the divergence between pedagogical and cognitively-plausible models of paradigmatic structure, and morphological description in general:

in his analogizing ... [t]he native user of the language ... operates in terms of all sorts of internally stored paradigms, many of them doubtless only partial; and he may first encounter a new basic verb in any of its inflected forms. (Hockett 1967: 221)

The intervening half century has seen the development of methodologies for probing speakers’ morphological knowledge, though the most robust measures, such as morphological family size (de Jong et al. 2000; Mulder et al. 2014), apply to derivational families. In parallel, large-scale corpora and other lexical resources have provided detailed information about the composition and structure of the forms ‘in circulation’ within a speech community. It is well established that forms exhibit a Zipfian distribution (Zipf 1935), in which the frequency of a word is (approximately) inversely proportional to its rank in a frequency table. There are two particularly significant consequences for the learning of inflectional systems. First, roughly half of the forms that a speaker encounters will be occurrences of a small number of high frequency items (*the*, *of*, etc., in English). Second, almost half of the items of a language will be hapax legomena, i.e., forms that occur exactly once in a corpus, irrespective of its size, and thus are unlikely to be encountered multiple times by speakers.

These distributional biases support Hockett’s conjecture about the nature of the input that a speaker encounters. They also argue against a naive pedagogically-influenced conception of paradigmatic structure and against any version of the ‘Full Listing Hypothesis’ (Hankamer 1989) on which speakers would be assumed to memorize the inflected forms of a language. Speakers appear to encounter only a small proportion of the forms of open-class items; for up to half of the lexicon, they may encounter just a single form. Nevertheless, speakers appear to

be able to ‘pool’ the variation exhibited by partial paradigms and extend attested patterns to new forms. The format of this knowledge and the mechanisms that speakers use to amalgamate and extrapolate patterns are not yet well understood. However, there is evidence to suggest that speakers exploit lexical neighborhoods to define an analogical base that can bootstrap the process (Blevins et al. 2017).

The morphological processing literature has also investigated a range of effects in which paradigmatic structure appears to play a significant role. Among the most striking of these is the **relative entropy** effects first reported in Milin et al. (2009a, 2009b). These studies found that speakers in a lexical decision task were sensitive to the divergence between the frequency distribution of an item’s inflectional paradigm and that of its inflection class. To take a specific example, the more closely the distribution of the forms of the Serbian noun *planina* ‘mountain’ matches the distribution of the ‘feminine’ declension in Serbian, the faster and more accurately speakers recognize forms of *planina*. This finding has been confirmed in follow-up studies including Baayen et al. (2011) and Milin et al. (2017).

3.3 Syntagmatic/paradigmatic integration

The interpretation of syntagmatic variation as, in part, a marker of paradigmatic relations also clarifies the motivation for the representational agnosticism of Neogrammarians such as Paul (1920). On the type of ‘constructive’ (Blevins 2006) interpretation of structure adopted in Post-Bloomfieldian models, forms are composed of a determinate sequence of atomic elements. On a more agnostic ‘abstractive’ interpretation, structure emerges from patterns exhibited by sets of forms. Multiple form classes are potentially relevant for the purposes of discriminating forms and deducing implications. Hence, the determination of syntagmatic structure, like the enumeration of inflection classes, is task-dependent. The structure exhibited by forms, like the number of classes contained by a language, depends on the purposes for which one is assigning structure or counting classes.

Current approaches to the modelling of inflectional patterns address a broad range of other issues. Network architectures, particularly ‘wide learning’ networks (Baayen and Hendrix 2017) provide a representation of paradigmatic structure that avoids the ‘combinatorial explosion’ (Baayen et al. 2013) associated with lists of forms, particularly if surface variants are encoded by exemplars. Models of vector semantics provide a way of grounding the interpretation of forms in an observable dimension of variation, i.e. distribution. The other observable dimension, variation in shape, is described by discriminable contrasts. Taken together, these dimensions offer an interpretation of the features

employed in the description of inflectional systems, without assigning a ‘meaning’ to the features themselves.

4 Periphrasis and grammaticalization

We have to do with periphrasis if for certain cells of an inflectional paradigm no synthetic morphological form is available. Instead, a combination of words, an analytic or periphrastic form has to be used. For instance, Latin has no synthetic forms for the perfective passive of verbs, as illustrated here for the 3SG forms of the verb *laudāre* ‘to praise’:

Table 1: Imperfective and perfective 3SG forms of *laudāre*.

Imperfective	Active	Passive
PRESENT	<i>laudat</i>	<i>laudātur</i>
PAST	<i>laudābat</i>	<i>laudābātur</i>
FUTURE	<i>laudābit</i>	<i>laudābitur</i>
Perfective	Active	Passive
PRESENT	<i>laudāvit</i>	<i>laudātus/a/um est</i>
PAST	<i>laudāverat</i>	<i>laudātus/a/um erat</i>
FUTURE	<i>laudāverit</i>	<i>laudātus/a/um erit</i>

The cells for the perfective passive are a combination of the passive participle (that, like adjectives, agrees with the subject of the clause with respect to case, gender and number) and a form of the verb *esse* ‘to be’. If these word combinations were not considered part of the verbal paradigm, Latin verbs would have a paradigm with a gap for the perfective passive forms. These periphrastic forms receive a perfect interpretation, although the forms of the verb *esse* ‘to be’ are that of the imperfect tense.

An additional argument for considering these word combinations as filling paradigm cells is the following. Latin has a number of so-called deponent verbs, verbs with a passive form but an active meaning. For instance, the verb *lo-quor* ‘speak’ is such a deponent verb. The crucial observation is that a word-sequence such as *locutus est* receives an active interpretation as well, and means ‘he has spoken’. This parallelism in interpretation as active meanings is to be expected if these analytic forms belong to the inflectional paradigm of verbs (Börjars et al. 1997).

This means that phrasal constructions may express inflectional properties, just like morphological constructions (Ackerman and Stump 2004; Börjars et al. 1997; Sadler and Spencer 2001). Moreover, they cannot be analyzed as regular syntactic combinations because the morphosyntactic features of these constructions cannot always be determined on the basis of those of their word components (Ackerman and Stump 2004; Popova 2010). This can be illustrated by the periphrastic constructions for perfect tense in Dutch. These are combinations of the verbs *hebben* ‘have’ or *zijn* ‘be’ with the past participles of verbs, as in:

- (2) *Jan heeft het boek ge-lez-en*
 John have-3S DET book PTCP-read-PTCP
 ‘John has read the book’
- (3) *Het meisje is ge-vall-en*
 DET girl be.3S PTCP-fall-PTCP
 ‘The girl has fallen’

These two verbs are called auxiliaries, as they do not have their regular meaning in periphrasis, but instead express perfect aspect. The auxiliaries in these sentences have a present tense form. The past participles do not carry the perfective meaning either, as they can also be used with an imperfect meaning in passive sentences. Hence, the perfect meaning is the holistic property of the word combination as a whole. This is why we speak of periphrastic constructions, because constructions are form-meaning combinations with possibly holistic properties that cannot be deduced from properties of their constituents.

Periphrastic constructions require a formal analysis which does justice to these holistic properties. One model is that of Paradigm Function Morphology. In this model phrasal combinations can function as the realization of the set of morphosyntactic and/or morphosemantic features of a lexeme. For instance, *heeft gelezen* in example (2) is treated as the realization of the lexeme *lezen* ‘to read’ with the feature values [3.sg.perf] (Popova 2010).

In the framework of Construction Morphology, periphrastic constructions are accounted for by means of the notion of ‘constructional idiom’ (Booij 2010). In English, for instance, the perfect tense form of verbs is a complex verbal predicate that consists of a form of the verb *have* with a past participle. This specific pattern expresses the perfect meaning, which cannot be derived from the meaning of one of the constituent words: neither the verb *have*, nor the participle itself is the carrier of the perfect meaning. The following constructional idiom may be assumed for the English perfect tense construction:

- (4) $[[have]_{Vi} [past\ participle]_{Vj}]_{Vk} \leftrightarrow [PERF [SEM_i]]_{SEM_k}$

where PERF stands for the meaning ‘perfect tense’, and SEM for the meaning of the relevant indexed constituent.

The use of auxiliaries for periphrastic constructions is a case of grammaticalization, i.e. the process by which words with an original lexical meaning acquire a more grammatical meaning (Hopper and Traugott 1993). For instance, English *have* has the lexical meaning of ‘possess’ but this is not the relevant meaning in the periphrastic tense forms. Dutch has two auxiliaries for perfect tense, reflecting two different historical sources of these constructions, a possessive construction and a predicative one. In English the predicative construction was the source of the use of the auxiliary *be*, as in older English *He is fallen* ‘He is in a situation of fallen-ness’, which is now replaced in present-day English by *He has fallen*. This illustrates that English *have* is even more grammaticalized than its Dutch counterpart.

5 Lexicon and inflection in computational morphology

In computational terms, many aspects of the theoretical debate dealt with in the previous sections boil down to the problem of dealing with cases of many-to-many surface relations between form and function at the word level. Consider the representations in (5) and (6), where the Italian verb forms *tengo* ‘(I) hold’ (present indicative, 1st person singular) and *tieni* ‘(you) hold’ (present indicative, 2nd person singular) – from the verb *TENERE* ‘hold’ – are segmented into their surface constituents.

- (5) $[[teng]_{\text{HOLD}} + [o]_{\text{pres ind 1s}}]_{\text{HOLD pres ind 1s}}$

- (6) $[[tien]_{\text{HOLD}} + [i]_{\text{pres ind 2s}}]_{\text{HOLD pres ind 2s}}$

The same word forms are split into more abstract sub-lexical constituents in (7) and (8):

- (7) $[[ten]_{\text{HOLD}} + [o]_{\text{pres ind 1s}}]_{\text{HOLD pres ind 1s}}$

- (8) $[[ten]_{\text{HOLD}} + [i]_{\text{pres ind 2s}}]_{\text{HOLD pres ind 2s}}$

Representations in (7) and (8) define a sign-based (biunique) relationship between the stem *ten* and its lexeme HOLD. This is blurred in (5) and (6), where lexical formatives are not identical, leaving us with the problem of classifying *teng* and *tien* as stem allomorphs of the same verb. In the generative linguistic literature, shorter lexicons are generally preferred over more verbose ones, under the standard generative assumption that what can be computed should not be stored. However, for a fully-inflected form to be produced from the representation in (7), some adjustment rules have to be added to the grammar (e.g. a velar insertion rule before back vowels in the case at hand). In the Italian conjugation, many of these rules do not apply across the board, but obtain for particular lexemes in specific contexts only (Pirrelli and Battista 2000). So the question of descriptive economy cannot really be confined to the lexicon, since a compact lexicon may call for a rather profligate set of *ad hoc* rules. Information theory provides a means to address these empirical issues on a principled basis. In an information-theoretic adaptation of Harris' ideas (1951), Goldsmith (2001, 2006) models the task of morphological induction as a data compression (or Minimum Description Length, MDL) problem: "find the battery of inflectional markers forming the shortest grammar that best fits the empirical evidence" (Rissanen 1989). The grammar is a set of paradigms, and the empirical evidence a reference corpus, where each form occurs with a specific frequency distribution. In this framework, MDL penalizes two descriptive extremes. First, it disfavors an extremely redundant grammar, where each word form is part of a singleton paradigm that has that form as its only member. This in fact amounts to a repository of fully listed words, where each form is assigned the probability with which it occurs in the corpus. At the opposite end, a very compact grammar contains one overall paradigm only, where any verb stem can freely combine with any affix. This is a very short and redundancy-free grammar, but it overgenerates wildly, thus providing a poor approximation of the distributional evidence of the forms attested in the corpus. In what follows, we consider a rather more classical computational approach to this issue, using Finite State Transducers.

5.1 Finite State Transducers

A Finite State Transducer (FST) is an abstract computational device that turns an input string into an output string. Figure 1 depicts an FST for the present indicative stems of the Italian verb *TENERE* 'hold': namely *ten-*, *tien-* and *teng-*. In the graph, nodes (circles) are memory states, and directed arcs (arrows) represent transitions from one state to another. Each arc is decorated with either a single

symbol, or a pair of symbols separated by a colon. The colon is a mapping operator and reads: “the symbol on the left (or lexical symbol) is mapped onto the symbol on the right (either a surface symbol or a gloss)”. If only one symbol appears on the arc, this means that the symbol is mapped onto itself, i.e. lexical and surface symbols are identical. Lower-case characters in the Latin alphabet represent simple letters; ‘ ϵ ’ is a meta-symbol representing the null character; and ‘ Σ ’ is a variable ranging over any input symbol. Finally, upper-case Latin characters are linguistic glosses that express morphosyntactic features (e.g. ‘PAST’ or ‘3S’), stem indexes (e.g. ‘ B_1 ’) and lexical content (‘TENERE’).

Figure 1: A Finite State Transducer for the present indicative stem allomorphs of Italian *TENERE* ‘hold’.

Starting from the initial state q_0 , the FST reads one input symbol at a time from left to right, until it consumes the whole input string. Upon reading a symbol (e.g. ‘t’ in *teng-*) in a particular state, the transducer looks for a state-leaving arc that is labelled with the same input symbol. If such an arc exists, it is traversed, and the mapping operation annotated on the arc is carried out. An input string is successfully processed if the FST finds itself in the final state f after reading the final letter of the input string. Note that all branches of the FST in Figure 1 end up with a stem index (B_i) being inserted: e.g. ‘ $\epsilon: +B_2$ ’. The index enforces the constraint that the just read off stem allomorph is followed by an inflectional ending selecting the appropriate index.¹

¹ Stem indexing (Stump 2001) enforces a relationship between a verb stem and a set of cells in the verb paradigm. Although stem indexes and stem formation processes are often mutually implied (as illustrated by the FST in Figure 1), they might occasionally be independent (Pirrelli & Battista 2000).

For example, in the present indicative, *teng-* can be followed only by the first singular and third plural ending.

A number of properties make FSTs theoretically interesting devices. FSTs are not classical rewrite rules. They enforce a correspondence relation between two distinct *levels* of representation. The relation is declarative, parallel and bi-directional. We can easily swap input and output symbols and use, for word generation, the same transducer designed for word parsing. This means that we can understand the *teng-ten* relation in either of the following ways: (i) *g* is trailed after *ten-* to incrementally (in Stump's 2001 terms) add B_2 to the resulting representation; (ii) *teng* is the result of applying to *ten* a specific paradigm function that realizes the appropriately indexed allomorph B_2 . In the former case B_2 adds some (disjunctive) morphological content. In the latter case, it is only the indexed trace of a realizational process.

Secondly, FSTs represent lexical items as a union set of intersecting automata. Since automata define processing operations over lexical and surface representations, the whole finite state machinery amounts to a two-level word graph incorporating lexical information, morphotactic constraints and context-sensitive formal changes, all cast into a uniform format. FSTs thus blur the distinction between lexical representations and rule schemata defined over lexical representations. This makes them extremely powerful processing devices, capturing morphological generalizations at fine-grained levels of generality, ranging from sweeping, exceptionless phonological processes to lexically-conditioned surface adjustments. The overall emerging view is closer to the idea of a large, integrated repository of both specific and general information, than to the classical notion of a generative grammar consisting of a lexical repository and a set of rules.

On a less positive note, the high expressive power of FSTs can get a major liability. For a lexicon of average inflectional complexity, it soon becomes extremely cumbersome to craft by hand the whole range of fine-grained mapping relations that are needed for recognition/production. And this can be uselessly laborious. For example, since every stem must be associated with a specific cell index, transitions to cell-specific indexes must also be stipulated for the singleton stem of a regular paradigm. We thus miss the obvious generalization that only irregular paradigms require explicit stipulation of the transitions from stem alternants to specific subsets of paradigm cells. The DATR formalism (Evans and Gazdar 1996) provides an effective framework for implementing default statements through inheritance relations in the lexicon, and can be used to enforce both high- and low-order constraints on stem allomorphy.

5.2 Hierarchical lexicons

A DATR lexicon is an inheritance network of nodes, each consisting of a set of morphologically related forms, ranging from major inflection classes (Table 2, columns 1 and 2), to regular (Table 2, column 3) and subregular paradigms (Table 2, column 4). An inflected form like, e.g., *tengo* is obtained from the abstract lexical schema ‘X + o’ by assigning ‘X’ an appropriate stem.

Each entry node is defined by a “name” (the unique identifier in bold ended by a colon), followed by a few “statements”, each taking one line in Table 2, with the following general format: “attribute == value” (where ‘==’ is an assignment operator). An attribute is surrounded by angle brackets. It can consist of a single label (e.g. “<B>”, in Table 2), or by a sequence of blank-separated labels, defining an increasingly specific attribute (e.g. “<B 2>”). The empty label ‘<>’ denotes the most general or unspecified attribute. Any label added to ‘<>’ defines a branching node in the hierarchy: the further down we go in the hierarchy, the more specific the attribute. Accordingly, ‘<pres ind 3S>’ is more specific than ‘<pres ind>’, which is in turn more specific than ‘<pres>’. A value can be a constant string (e.g. *tem* in column 3, Table 2), or another attribute, either simple or complex.

Table 2: DATR lexical entries for Second Conjugation (2C) Italian verbs.

1	2	3	4
Morpholex-paradigm-0:	2C_VERB:	TEMERE:	TENERE:
<B>== “<root>”.	<>== Morpholex-paradigm-0	<>== 2C_VERB	<>== 2C_VERB
	<pres ind 1S>== “<B 2>” o	<root>== t e m.	<B 2>== t e n g
	<pres ind 2S>== “<B 3>” i		<B 3>== t i e n
	<pres ind 3S>== “<B 3>” e		<root>== t e n.
	<pres ind 1P>== “<B 4>” i a m o		
	<pres ind 2P>== “<B>” e t e		
	<pres ind 3P>== “<B 2>” o n o.		

Hierarchies of specificity play a crucial role in DATR information flow. General information, defined by attributes higher up in the hierarchy, tends by default to percolate to lower (more specific) attributes, unless the latter explicitly contain overriding information. For example, the node ‘Morpholex-paradigm-0’ (column 1, Table 2) states that the index ‘B’ is assigned whatever string is assigned to the ‘root’. This information is inherited by the node ‘2C_VERB’ (short for “second conjugation verb”: column 2, Table 2) through the statement “<>== Morpholex_paradigm_0”, which reads: “whatever

information is in `Morpholex_paradigm_0` is copied here". This statement boils down to adding all statements in 'Morpholex_paradigm_0' to '2C_VERB'. At this level, however, 'root' is assigned no surface string. This is done in the node 'TEMERE' (column 3, Table 2), which stipulates that all statements of '2C_VERB' ("`<>== 2C_VERB`") are inherited locally: here, the attribute 'root' is instantiated with the string *tem*. In turn, 'B' is assigned the local value of 'root' ("`<B>= <root>`"). At the same time, more specific 'B' attributes ('B 2', 'B 3', and 'B 4') inherit the string *tem* as their value, for want of more specific information being assigned locally. Inheritance thus captures the general statement that, in regular verbs, the basic stem 'B' is assigned by default to all present indicative cells. Conversely, in an irregular verb like 'TENERE' (column 4, Table 2), stem indexes ('B 2' and 'B 3') are explicitly assigned specific allomorphs (respectively *teng* and *tien*), thereby overriding the default distribution of the string *ten* canonically assigned to 'root'.

Thanks to default inheritance, DATR can describe fragments of comparatively complex inflectional systems like the Italian conjugation in a compact and elegant way (Pirrelli and Battista 2003). Note that fully-specified paradigms and abstract paradigmatic schemata are represented with the same toolkit of formal tools, in line with the view that they are statements of the same kind, which differ only in coverage. These statements are expressed in terms of paradigmatic relations between stem indexes, and bear witness to the theoretical usefulness of word paradigms as descriptive formal devices (Pirrelli 2000; Blevins 2003, 2006, 2016). We will return to issues of inter-cell predictability later in this chapter (Section 8), in connection with the problem of measuring the inflectional complexity of a language in information theoretic terms (e.g. Ackerman and Malouf 2013).

DATR offers a handy computational framework for modelling Jackendoff's (2002) idea of the lexicon as containing entries whose information may range from very general (i.e. obtaining for an entire class of verbs) to very specific (i.e. holding for one verb lemma or one verb form only). In DATR, the lexicon is not just a list of stored units, but defines the linguistic domain where morphological processes apply, thus coming very close to a rigorous computational framework for testing the Lexicalist hypothesis (Halle 1973; Jackendoff 1975; Aronoff 1976; Scalise 1984; Lieber 1992). Someone may keep considering it useful and conceptually desirable to draw a line between pieces of lexical information that are actually listed (and thus form the lexicon in a strict sense) from those which are computed on-line through general morphological statements. Nonetheless it soon gets very difficult to comply with this principled distinction, especially when it comes to the description of inflectional systems of average complexity (see Corbett and Fraser 1993 for another example with Russian inflection).

Upon reflection, this is not just an issue of descriptive accuracy. In fact, it boils down to the deeper, explanatory question of how children can come up with the relevant knowledge needed to optimally process inflected forms. According to this learning-based view, redundant patterns are predominantly statistical, and even irregularities appear to be motivated by their frequency distribution in the system and the general-purpose learning strategies of the human language processor. All of this can admittedly be very different in character from the formal constraints on units, representations or rule systems proposed within theoretical and computational models. Nonetheless, it represents, in our view, an extremely insightful entry point to the grand issue of language architecture. In the following section we shortly consider some implications of the regular vs. irregular distinction from a developmental perspective, to then move on to the computational modelling of inflection acquisition (see Ravid, Keuleers and Dressler 2020, this volume, for a more comprehensive overview of morphology acquisition).

6 Acquiring inflection

6.1 The logical problem of acquiring inflection

According to a classic account (Berwick 1985; Pinker 1989, 1984), the task of a child attempting to learn how verbs and nouns are inflected can be described as involving grammar hypothesis testing. This is not too dissimilar from the MDL grammar evaluation framework illustrated in the previous section. Children need to arrive at a grammar hypothesis *H* that includes all well-formed inflected forms of the input language, while excluding ill-formed ones. An important logical problem with this account is represented by the case when the child's grammar is a superset of the target grammar (Pinker 1989): the child masters all correctly inflected forms, but nonetheless also produces some ungrammatical forms (e.g. (s)he says *went* and **goed* interchangeably). How can children recover from these errors?

In the acquisitional literature, the question of whether children can correct themselves on the basis of received explicit negative evidence (e.g. negative correction by their care-givers) has been highly debated. Some scholars suggest there is little reason to believe that children are supervised in this way (Bruck and Ceci 1999; Taatgen and Anderson 2002; but see Chouinard and Clark 2003 for a different view). Even when they are corrected (which is neither frequent nor systematic), they may take little notice of correction. In contrast with this position, other scholars (Kilani-Schoch et al. 2009; Xanthos et al. 2011)

emphasize the impact on child's morphology learning of both explicit and implicit, positive and negative feedback by parents, thereby questioning speculative claims (of direct or indirect Chomskyan inspiration) of poor and noisy input evidence in child-directed speech.

Proponents of lack of corrective feedback in child's input have hypothesized that possibly innate mechanisms for self-correction are in place. For example, according to Marcus et al. (1992), a *blocking mechanism* may suppress *ed*-verb past formation in English *goed, due to the stored representation of irregular *went* being entrenched in the lexicon by repeated exposure. This is in line with Pinker and Ullman's (2002) 'Words and Rules' theory, according to which only irregularly inflected forms are stored in full. Regulars are either combined online from their stems and affixes in word generation, or are split into stems and affixes in recognition under the assumption that their access units are sublexical. In producing an inflected form, the lexicon is accessed first, and on-line assembly is pre-empted if the target (irregular) form is found there. Otherwise, a regular form is produced by using combinatorial rules.

However logically elegant and simple, lexical blocking seems to make rather unrealistic assumptions about how children come up with a regular vs. irregular distinction while being exposed to inflectional systems of average complexity. For example, Pinker and Ullman (2002) suggest that there is only one default regular pattern in the inflection of any language, and that the decision is dichotomous: the child considerably constrains the hypothesis space, as only one inflectional process can be held as a "regular" candidate for any specific paradigm function. This assumption, however, appears to seriously underestimate the systemic complexity of highly-inflecting languages. For example, the verb conjugations of Italian and Modern Greek show a graded hierarchy of regularity-by-transparency effects of morphological processing: many verbs present phonologically predictable adjustments that obscure the stem-ending boundary; some others keep the transparency of the stem-ending boundary at the price of introducing formally unpredictable fillers like thematic vowels, etc. Thus, the central question is how children can possibly home in on the decision of storing an irregular form as an unsegmented access unit, and a regular form as consisting of at least two distinct access units. We know that, for some models of lexical access, the decision does not have to be yes-or-no. For example, so-called "race models" (Schreuder and Baayen 1995; Baayen et al. 1997) assume that words can possibly be stored as both whole forms and sublexical access units, thus providing two parallel access routes that are concurrently activated and adjudicated on the basis of frequency and task-based effects. However, the same models leave the issue of word segmentation seriously underspecified: how does the child segment a morphologically complex word

that is taken to be regular? How and when does perception of sublexical structure develop in child lexical competence to trigger lexical self-organization?

In a comprehensive comparison of the developmental stages in the acquisition of verb inflection in nearly two dozen languages (the Indo-European, Ugro-Finnic and Semitic families plus Turkish), Bittner, Dressler and Kilani-Schoch (2003) observe that the transition from rote lexical storage to morphological processing is the result of a process of active knowledge construction by the child, crucially conditioned by typological factors such as richness, uniformity and transparency of inflectional paradigms (Dressler 2010). Nevertheless, scholars widely differ in the way they conceptualize this process.

In the framework of Natural Morphology (Dressler 2010, 2005), an increase in children's inflectional productivity follows the establishment of the first "mini-paradigms": i.e. non-isolated mini-sets of minimally three phonologically unambiguous and distinct inflected forms of the same lemma. Miniparadigms thus mark the turning point between a "premorphological" and a "protomorphological" phase in child acquisition of inflection. The onset of this transition phase can vary considerably, depending on type and token frequency, lexicon size, phonological salience, regularity and transparency of the morphological system.

Legate and Yang (2007) model child acquisition of English inflection as a maturation period during which the child starts entertaining a grammar hypothesis compatible with a language that does not manifest tense marking (e.g. Chinese), to eventually eliminate it. The crucial observation here is that elimination of this hypothesis is a gradual process. The frequency of inflectionally unmarked usages by the child goes down incrementally, mostly over a 2–3-year span (Haegeman 1995; Phillips 1995). More interestingly for our present concerns, the length of such a maturational period and the frequency of inflectionally unmarked usages are influenced by the richness and complexity of the inflection system being acquired. Based on distributional evidence in the child-directed portion of the Brown's (1973) Harvard Studies, the Leveille corpus and the Geneva corpus in the CHILDES database (MacWhinney 1995), Legate and Yang observe that the percentage difference between clauses with overt tense morphology (e.g. *I walked to school*) and clauses with no tense morphology (e.g. *I make him run*) is nearly 6% in English, 40% in French, and 60% in Spanish. Accordingly, they make the prediction that the maturational period for a child to acquire contextually appropriate tense marking will be longer for English and shorter for Spanish. This is borne out by developmental data. The greater balance of inflectional contrast in the Spanish verb system shows an inverse correlation with the lower percentage of unmarked usages observed in Spanish children (17%), compared with German children (58%) and English children (87%) (data from the CHILDES corpus, reported in Freudenthal et al. 2010).

In a series of influential papers on discriminative language learning, Ramscar and colleagues (Ramscar and Yarlett 2007, Ramscar and Dye 2011) focus on the logical problem of acquisition of English noun inflection to show that children can in fact recover from superset inflection grammars by learning probabilistically-cued responses, according to the Rescorla-Wagner model of classical conditioning (Rescorla and Wagner 1972). Early in training, children go through a transitional phase where more forms like – say – **mouses* and *mice* are produced interchangeably. Due to the high number of -s plural nouns, compared with the relatively small frequency of *mice*, the pressure for *mice* to be replaced by the over-regularized **mouses* is strong and dominates competition. In the end, however, children’s over-regularized response is surpassed by the “imitation” signal, i.e. by the increasing strength of the imitative response *mice* when its performance curve reaches asymptote.

The discriminative account dispenses with both innate blocking mechanisms and parameterized grammar hypotheses. In the acquisition of verb inflection, input evidence of tense-marked forms like *walked*, *ate* and *sang* strengthens the association with the implicit morpho-semantic notion of [+Past] and the referential content of the three forms (namely WALK, EAT and SING). Root infinitives of the same verbs, on the other hand, will in turn reinforce associative links with [– Past], and, over again, with WALK, EAT and SING. Thus, at early stages of English acquisition, the lexical semantics of each verb is more likely to be associated with verb forms that are unmarked for tense than with tensed forms (in line with Legate and Yang’s distributional evidence). This intra-paradigmatic competition (i.e. competition between finite and nonfinite forms belonging to the same verb paradigm) turns out to be more balanced in a richer and more complex inflection system, like Spanish or Italian conjugation, where paradigm cells are, in the vast majority of cases, associated with formally distinct, fully contrastive verb forms.

6.2 Paradigms and entropy

It is useful to see how word competition is related to the entropy $H(P)$ of the distribution of phonologically distinct members in the verb paradigm P , or “contrastive” paradigm entropy, defined in equation (1):

$$H(P) = - \sum_{f_i \in P} p(f_i) \log_2(p(f_i)) \quad (1)$$

where f_i ranges over all formally distinct inflected forms of P , and the probability $p(f_i)$ is estimated by the ratio between the token frequency $t(f_i)$ and P cumulative frequency $\sum_{f_i \in P} t(f_i)$.

Table 3 shows a (fictitious) distribution of inflected verb forms in three present indicative (sub)paradigms for English *sing*, German *singen* ‘sing’, and Spanish *cantar* ‘sing’, under the simplifying assumption that distributions do not vary across languages, but only depend on lexical information (the verb being inflected) and morphosyntactic information (the selected paradigm-cell). When we calculate the entropy of each present indicative paradigm based on its distinct inflected forms, we are assessing how uniformly distributed the formal contrast is within the paradigm. Even if we assume that frequency distributions do not vary across languages, the comparative complexity² of the inflection system in each language appears to affect the amount of intra-paradigmatic formal competition: the more discriminable the inflected forms are, the more entropic (i.e. the more balanced) their competition will be. When more cells are assigned the same form, cumulative frequency by form winnows down contrastive paradigm entropy, prompting a greater bias for fewer forms. Accordingly, the English bare form *sing* is largely dominant in its own present indicative paradigm.³

Table 3: Frequency distributions of verb forms in the present indicative paradigms of English, Spanish and German; $p(pc|s)$ is the conditional probability of a paradigm cell pc given a verb stem s .

pres ind	English			Spanish			German		
p-cell	sing	freq	$p(pc s)$	cantar	freq	$p(pc s)$	singen	freq	$p(pc s)$
1S	<i>sing</i>	5	0.14	<i>canto</i>	5	0.14	<i>singe</i>	5	0.14
2S	<i>sing</i>	2	0.06	<i>cantas</i>	2	0.06	<i>singst</i>	2	0.06
3S	<i>sings</i>	10	0.29	<i>canta</i>	10	0.29	<i>singt</i>	10	0.29
1P	<i>sing</i>	4	0.11	<i>cantamos</i>	4	0.11	<i>singen</i>	4	0.11
2P	<i>sing</i>	2	0.06	<i>cantáis</i>	2	0.06	<i>singt</i>	2	0.06
3P	<i>sing</i>	12	0.34	<i>cantan</i>	12	0.34	<i>singen</i>	12	0.34
	Σ	35	1	Σ	35	1	Σ	35	1

² The term “complexity” is used here in the intuitive sense of “formal variety/richness”. We will provide a more rigorous definition in Section 8 of this chapter.

³ In the case of the *sing* present indicative subparadigm, contrastive entropy is calculated as follows:

$$-p(sing) \cdot \log_2(p(sing)) - p(sings) \cdot \log_2(p(sings)) = 0.87,$$

where $p(sing)$ and $p(sings)$ are estimated as the ratio between their respective token frequencies in the subparadigm (25 and 10), and their cumulative token frequency (35). For the German and Spanish subparadigms, the same formula yields higher entropy values, respectively 1.68 and 2.28.

An important difference between Legate-Yang's model and the naive discriminative approach of Ramscar and colleagues is that, unlike the latter, the former makes the finite vs. non-finite competition hardly sensitive to lexical patterning in the data: any usage of finite marking in the child's input will reward a [+Tense] grammar *across the board*, lending support to any prospective tense-marking usage, irrespective of whether it involves an auxiliary, a copula or a lexical verb. Legate and Yang's model assumes that children are not learning how to inflect words, but how to reject a particular grammar hypothesis (or parameter setting). Contrariwise, a discriminative model makes any further usage of tense marking contingent upon the amount of contrastive competition *within* a specific paradigm.

The prediction that paradigm-specific distributions of finite and nonfinite forms play a role in determining the time course of inflection development is supported by the pattern of results reported in four quantitative analyses of early child language (Hamann and Plunkett 1998; Krajewski et al. 2012; Pine et al. 2008; Wilson 2003). In particular, by counting the number of times three classes of inflected verb forms (1SG and 3SG copula BE, 1SG and 3SG auxiliary BE, and 3SG -s) are overtly realized by English speaking children in obligatory contexts relative to all such contexts (or production rate), a few key findings are reported. First, there are significant differences in the rate at which typically developing children produce copula BE (e.g. *It's good*), auxiliary BE (e.g. *I'm eating*) and third person singular present forms (e.g. *He runs*), as indicated by the following ranking: *cop BE* > *aux BE* > *3SG*. The second finding is that *is* is produced at a higher rate with pronominal subjects than with lexical subjects. Thirdly, both copula and auxiliary *is* are produced at a significantly higher rate than *am*; among all instances of *is*, those preceded by *it* are realized significantly more often than those preceded by *he*.

6.3 Form and distribution effects

Discriminative learning paves the way to a coherent operationalization of inflection acquisition in terms of learning contrasts within and across the two observable dimensions of variation in an inflectional system (Blevins 2016): form classes and frequency distribution classes. The discovery of phonological and semantic sublexical invariants (e.g. stems in lexical paradigms, or inflectional endings in conjugation classes), has often been invoked as a mechanism accounting for the acquisition of inflectional morphology (Bybee 1988, 1995; Peters 1997; Penke 2012). Furthermore, an explanatory mechanism that crucially rests on the amount of competing sources of discriminative information in the input can naturally be extended beyond the word level. To illustrate, the following nuclear sentences form a

sort of *multi-word* or *syntactic paradigm*,⁴ defined as a set of *constructions* in complementary distribution and mutual competition.

- (9) *She walks to school*
She drinks cola
She's walking to school
Does she drink cola?

From this perspective, “[g]rammatical features, properties and categories can likewise be interpreted as proxies for form classes, distribution classes or some combination of the two. In this way, morphological terminology that misleadingly implies an associated semantics can be reduced to robustly observable dimensions of form variation” (Blevins 2016: 249). *Prima facie* sign-based relationships between form and content like 3SG -s in English verb conjugation are derived from the relation between a morphological marker and its embedding syntactic context. From this perspective, inflectional paradigms are descriptively useful shorthands for construction-based paradigms, where discriminative relations are expressed between fully spelled-out lexical forms and grammatical forms, rather than between word forms and abstract paradigm cells.

The idea is in line with Harris’s (1968) distributional approach to word segmentation and morpheme segmentation, where structural boundaries are identified at the points of likelihood discontinuity between adjacent symbols. We will consider these issues in more detail in Section 7.2.2. Note that a distributional, discriminative approach to inflection in fact conflates the two interlocked issues of how inflection is marked and in what contexts (see Section 1 above). Both aspects appear to capture distributional relations, and can be viewed as the same linguistic phenomenon looked at on different time scales. The distributional constraints that require verb marking, or agreement between two co-occurring units, are captured within a relatively large temporal window. Realizational effects are perceived on a shorter time scale. Contexts, as well as complex words, are split at their “joints” by contrasting “minimal” pairs of neighbors. Accordingly, paradigmatic relations emerge as contrastive points in a multidimensional space where shared lexical units (e.g. *she WALKs* vs. *he WALKed*) or shared grammatical units (e.g. *SHE IS walking* vs. *SHE IS singING*) are observed to occur in complementary distribution.

---

⁴ We can trace back the first formulation of this idea to P.H. Matthews’ *Syntax* (1981: 265–291), where a prescient construction-based interpretation of Chomskyan transformational rules is offered.

This view is supported by evidence that children acquire inflected forms by binding them to larger unanalyzed word “chunks” (Cazden 1968; MacWhinney 1976). Further support comes from evidence of children with Specific Language Impairment⁵ (SLI), who have difficulty with the interpretation of a complex sentence like *The cow sees the horse eating* (Leonard and Deevy 2011; Leonard et al. 2013). In addition, Purdy et al. (2014) show that older children with a history of SLI present a P600⁶ for subject-verb agreement violations as in **Every night they talkS on the phone*, but not for violations as in **He makes the quiet boy talkS a little louder*. If the children fail to grasp the structural dependencies between *He makes* and the clause that followed, no error would be detected, because the sequence *the quiet boy talks a little louder* would seem grammatical.

This data makes an interesting connection with the most significant processing limitations observed in children with SLI: slow processing speed (Stark and Montgomery 1995; Kail and Leonard 1986; Wulfeck and Bates 1995, among others) and limited phonological working memory capacity (Briscoe et al. 2001; Farmer 2000; Montgomery 2004, among others). In particular, if children with SLI are weak in retaining the phonological representation of a new word (for long enough for it to be stored in long term memory), this may result in a limited lexicon and a limited grammar. Development in the processing and production of complex sentences may also be delayed as a consequence of an impoverished lexicon, since complex sentences often involve verbs subcategorizing for sentential arguments. A more direct consequence of a limited working memory capacity in children with SLI is the difficulty to retain distant dependency relations in context, which explains their apparent insensitivity to violations as in **He makes the quiet boy talkS a little louder*.

To sum up, inflection appears to develop in children as a highly interactive system, interfacing formal and functional features on different time scales. A rich, contrastive morphological system is helpful to acquire syntactic dependencies. At the same time, full mastery of inflection is contingent upon the intake of larger and larger syntactic contexts, where functional dependencies are realized as extended, multi-word exponents. Such a two-way implication is hardly

---

⁵ Specific Language Impairment is a significant deficit in language ability that cannot be attributed to hearing loss, low non-verbal intelligence, or neurological damage (Leonard 2014; Montgomery & Leonard 1998; Rice et al. 2000).

⁶ P600 is a peak in electrical brain activity that is measured with electroencephalography around 600 milliseconds after the stimulus that elicits it. P600 is commonly associated with hearing or reading grammatical (in particular, syntactic) errors (see Marangolo & Papagno 2020, this volume).

surprising upon reflection, since different forms of the same verb paradigm are less functional if no verb arguments are yet expressed in context (Gillis 2003).

We emphasized the largely distributional character of inflectional paradigms, whose cells define grammatical abstractions over a multi-dimensional network of formal contrasts in syntactic and pragmatic contexts. Nonetheless, paradigms acquire an autonomous relevance in word processing: they stake out the linguistic space where lexical forms get co-activated and compete in word recognition and production through contrastive formal oppositions.

Contrastive paradigm entropy correlates with inflection rates in child production, and marks an important typological difference between inflecting languages like Spanish or Italian and languages of the nearly isolating type such as English. Finally, as we will see in more detail in the ensuing sections, issues of intra-paradigmatic formal contrast are also relevant for understanding the communicative function of the opposition between regularly and irregularly inflected words.

7 Machine learning of inflection

The task of modelling, with a computer, the dynamic process whereby a child gets to acquire her/his full morphological competence is reminiscent of Zellig Harris' empiricist goal of developing linguistic analyses on the basis of purely formal, algorithmic manipulations of raw input data: so called “discovery procedures” (Harris 1951). Absence of classificatory information (e.g. morphosyntactic or lexical information) in the training data qualifies the discovery algorithm as *unsupervised*. Conversely, when input word forms are associated with output information of some kind, then discovery is said to be *supervised*, and the task is modelled as a classification problem.

7.1 Constructive vs. abstractive approaches

Borrowing Blevins' (2006) terminology, a useful distinction can be made here between *constructive* and *abstractive* algorithms for word learning. Constructive algorithms assume that classificatory information is morpheme-based. Word forms are segmented into morphemes for training, and a classifier must learn to apply morpheme segmentation to novel forms after training. An abstractive learning algorithm, on the other hand, sees morphological structure as emerging from full forms, be they annotated with classificatory information (supervised

mode) or not (unsupervised mode). From this perspective, training data consist of unsegmented word forms (strings of either letters or sounds), possibly coupled with their lexical and morphosyntactic content. Accordingly, morphological learning boils down to acquiring knowledge from lexical representations in training, to generalize it to unknown forms. In this process, word-internal constituents can possibly emerge, either as a result of the formal redundancy of raw input data (unsupervised mode), or as a by-product of form-content mappings (supervised mode).

In fact, for too many languages, morpheme segmentation is not a well-defined task, due to the notorious problems with the Bloomfieldian, sign-based notion of morpheme, and the non-segmental processes of introflexive (i.e. root and pattern), tonal and apophony-based morphologies. So, the assumption that any word form can uniquely and consistently be segmented into morpheme-like constituents is at best dubious, and cannot be entertained as a general bootstrapping hypothesis for morphology learning.

In some abstractive algorithms, discovery procedures for morphological structure are constrained by a-priori assumptions about the morphology of the language to be learned. For example, knowledge that the target language morphology is concatenative biases the algorithm hypothesis search for stem-ending patterns. Thus, although no explicit morpheme segmentation is provided in training, the way word forms are tentatively split into internal constituents presupposes considerable information about boundary relations between such constituents (e.g. Goldsmith 2001). In some other algorithms, an alignment between morphologically related forms is enforced by either (i) using fixed-length positional templates (e.g. Keuleers and Daelemans 2007; Plunkett and Juola 1999), or (ii) tying individual symbols (letters or sounds) to specific positions in the input representation (so-called “conjunctive” coding: Coltheart et al. 2001; Harm and Seidenberg 1999; McClelland and Rumelhart 1981; Perry et al. 2007; Plaut et al. 1996), or (iii) resorting to some language-specific alignment algorithms (Albright 2002) or head-and-tail splitting procedures (Pirrelli and Yvon 1999). However, the ability to recognize position-independent patterns in symbolic time series, like the word *book* in *handbook*, or the Arabic verb root shared by *kataba* ‘he wrote’ and *yaktubu* ‘he writes’, appears to lie at the heart of human learning of inflection. Hence, a principled algorithm for morphological bootstrapping should be endowed with a capacity to adapt itself to the morphological structure of the target language, rather than with a language-specific bias.

In “features and classes” approaches (De Pauw and Wagacha 2007; McNamee and Mayfield 2007), a word form is represented as a set of redundantly specified *n*-grams, i.e. possibly overlapping substrings of *n* characters making up the input string: for example, ‘*wa*’, ‘*al*’, and ‘*lk*’ for the string *walk*. *N*-grams have no

internal structure and may be order-independent. The algorithm may start with the hypothesis that each word form is in a class of its own, and uses a stochastic classifier to calculate the conditional probability of having a certain class (a word form) given the set of distributed n -grams associated with the class. N -grams that occur in many words will be poorly discriminative, whereas features that happen to be repeatedly associated with a few word forms only will be given a morphologically meaningful interpretation.

Discriminative approaches to learning such as “features and classes” have a lot to offer. First, they are able to deal with the problem of learning “a-morphous” morphologies on a principled basis, addressing traditional conundrums in the morpheme-based literature such as morphemes with no meanings (“empty morphemes”), meanings with no morphemes (“zero morphemes”), bracketing paradoxes, etc. Secondly, they seem to exploit the time-honored linguistic *principle of contrast* (Clark 1987, 1990), according to which *any* formal opposition can be used to mark a grammatical or lexical opposition. Thirdly, they bring word learning down to more general mathematical models of classical conditioning (Rescorla and Wagner 1972) in behavioral psychology, according to which cues are constantly in competition for their predictive value for a given outcome. In what follows we will focus on abstractive, discriminative approaches to word learning and explore their implications for models of inflection. Such approaches see inflectional morphologies as **complex adaptive systems**, whose internal organization is the dynamic, continuously changing outcome of the interaction of distributional properties of input data, levels of lexical representation, and innate learning and processing constraints.

7.2 Associative vs. discriminative approaches

It is useful to describe discriminative learning by contrasting it with classical models of associative learning. The *gradient descent* training of connections from input nodes to output nodes in a two-layer perceptron for word production is a well-known example of associative learning. In learning inflection, the task is to map an input representation (say “go PAST”) onto the corresponding output representation (*went*). With PDP connectionist networks (Rumelhart and McClelland 1986), input representations are “wired in” on the input layer through dedicated nodes. A letter string like #go# (where ‘#’ marks the start and the end of the string) is simultaneously input through context-sensitive, conjunctive encoding of each symbol together with its embedding context. Accordingly, the input g in #go# is encoded as a #_g_o node. Output representations are learned by adjusting connection weights through *back-propagation* of the error feedback. Back-propagation consists

in altering the weights of connections emanating from the activated input node(s), for the level of activation of output nodes to be attuned to the expected output. According to the “delta rule” (equation (2)), connections between the j^{th} input node and the i^{th} output node are in fact changed in proportion to the difference between the target activation value $\hat{h}_i$ of the i^{th} output node and the actually observed output value h_i :

$$\Delta w_{i,j} = \gamma \cdot (\hat{h}_i - h_i) \cdot x_j \quad (2)$$

where $w_{i,j}$ is the weight on the connection from the j^{th} input node to the i^{th} output node, γ is the network learning rate, and x_j is the activation of the j^{th} input node. Note that, for $x_j = 0$, the resulting $\Delta w_{i,j}$ is null. In other words, nothing changes in the connections emanating from an input node if that node is not activated.

The “delta rule” in equation (2) is primarily *associative*. Learning proceeds by increasingly associating co-occurring cues (input nodes) and outcomes (output nodes), or by dissociating them when explicit evidence to the contrary is provided by means of correction. Activation of an output node in the absence of an activated input node does not affect the connection from the latter to the former, and nothing is learned about their cue-response relationship. Interposition of a hidden layer of nodes mediating input and output nodes makes the relationship between input and output representations non-linear, but does not make it up for lack of explicit negative evidence. In the end, learning is based on the fundamental assumption that the network is systematically “corrected”, i.e. it is told that an input-output connection is right or wrong.

A more realistic connectionist approach to language learning is offered by Recurrent Neural Networks (hereafter RNNs), which provide a principled solution to the problem of learning words with no external feedback (Elman 1990; Jordan 1997). RNNs are multi-layer perceptrons equipped with a first layer of hidden nodes interposed between the input layer and the output layer, and an extra layer of hidden nodes containing a copy of the activation state of the hidden layer at the previous time tick. In RNNs, a string is input as a time series of symbols (not as a synchronous activation pattern), with each symbol being presented at a discrete time tick. Furthermore, word learning is conceptualized as a prediction-driven task. Upon presentation of an individual symbol on the input layer, an RNN must guess the upcoming symbol on the output layer. In a nutshell, the network learns to predict sequences of symbols based on its past experience. This is fairly ecological: prediction is known to be heavily involved in human language processing (Altmann and Kamide 2007; DeLong, Urbach, and Kutas 2005; Pickering and Garrod 2007). In addition, it provides a natural

answer to the lack-of-feedback problem. Everything the network has to do is to wait for an upcoming symbol to show up. If the symbol does not match the current network prediction, connection weights are adjusted for the current input symbol to be the most likely network's response when a similar sequence is presented over again.

Simple RNNs are serial memories. They can improve on sequence prediction by keeping memory of the immediately preceding context through patterns of re-entrant connections in the hidden layer. At the same time, they plausibly address the problem of context-sensitive encoding of input symbols. The way a symbol is internally encoded by a *RNN* crucially depends on the network's memory of embedding sequences where the symbol was found. Different training data will yield different internal representations. There is no need for pre-wired input nodes encoding specific symbols in specific contexts. In principle, a *RNN* trained on strings from one language, can be trained on strings from another language, gradually adjusting its input nodes for them to be able to capture different serial dependencies in the way symbols are distributed.

Simple *RNNs* are trained with the same delta rule used for simple perceptrons. They learn to make context-dependent predictions that approximate the conditional probabilities of ensuing elements. To illustrate, the conditional probability of having 'b' immediately following 'a' (or $p(b|a)$) is estimated by $p(a,b)/p(a)$. Since the connection between an input node representing 'a' and an output node representing 'b' is strengthened by equation 2 in proportion to how often the bigram *ab* is found in input (and the learning rate γ), the network's prediction of 'b' given 'a' will be a direct function of $p(a,b)$. However, due to equation (2), the network learns nothing about – say – $p(c,b)$ or $p(a,c)$, from the distribution of *ab*'s. The assumption is in line with a purely associative view of error-driven learning and is supported by the intuitive observation that serial predictions can rely on already processed symbols only.

Discriminative learning couples equation (2) with the idea that multiple cues are constantly in competition for their predictive value for a given outcome. Accordingly, learning proceeds not by simply *associating* co-occurring cues and outcomes, but by *discriminating* between multiple cues. In Rescorla-Wagner's (1972) equations, the associative step is complemented with a competition-driven step, which forces the associative strength to take a step down when the outcome is present and the cue is not. Recent applications of Rescorla-Wagner equations to modelling a number of language tasks (see Pirrelli et al. 2020, this volume) are based on estimating the connection weights between two levels of representation: one for raw strings, based on *n*-gram encoding, the other one for the symbolic encoding of morpho-lexical and morphosyntactic units.

Temporal Self-Organizing Maps (*TSOMs*) have recently been proposed to memorize symbolic time-series as chains of specialized processing nodes, that selectively fire when specific symbols are input in specific temporal contexts (Ferro et al. 2011; Marzi et al. 2014; Pirrelli et al. 2015). *TSOMs* consist of:

1. one layer of input nodes (where an input stimulus is encoded),
2. a bidimensional grid of processing nodes (the map proper),
3. two levels of independently trainable connections:
  - a. input connections
  - b. temporal connections.

Through the level of input connections, information flows from the input layer to all map nodes (one-way input connections). A second level of connectivity goes from each map node to any other node on the map (including itself), forming a pool of re-entrant temporal connections that update each map node with the state of activation of the map at the previous time tick (one-time delay). Like with *RNNs*, a word is input to a *TSOM* as a time series of symbols, one symbol at a time. At each time tick, activation spreads through both input and temporal connections to yield an overall state of node activation, or Map Activation Pattern, at time t . In particular we use $MAP_t(s)$ to refer to the Map Activation Pattern relative to the current stimulus s . The node with the top-most activation level in $MAP_t(s)$ is referred to as Best Matching Unit, or $BMU_t(s)$. Finally, a time series of sequentially activated *BMUs*, namely $\langle BMU_1(s_1), \dots, BMU_t(s_t) \rangle$ is called a *BMU* chain.

A full description of the *TSOM* architecture and learning equations can be found in Pirrelli et al. 2020, this volume. Suffice it to say here that the learning algorithm modulates weights on re-entrant temporal connections for *BMU* chains to optimally process the most likely input strings. In particular, for any input bigram ‘ ab ’, the connection strength between $BMU_{t-1}(a)$ and $BMU_t(b)$ will:

- (i) increase every time ‘ a ’ precedes ‘ b ’ in training (entrenchment)
- (ii) decrease every time ‘ b ’ is preceded by a symbol other than ‘ a ’ (competition and inhibition).

Steps (i) and (ii) incrementally enforce node specialization. The map tends to allocate maximally distinct nodes for the processing of (sub)strings, as a function of their frequency in training. To illustrate the effects of node specialization vs. sharing on the representation of inflected forms of the same paradigm, Figure 2 sketches two possible end states in the allocation of *BMU* chains responding to four German forms of *beginnen* ‘begin’: *BEGINNEN* (infinitive, 1p and 3p present indicative), *BEGINNT* (3s and 2p present indicative), *BEGANNT* (2p preterite) and *BEGONNEN* (past participle). In the left panel, *BMUs* are

arranged in a word tree. At any node n_i , one can always retrace backwards the nodes activated to arrive at n_i . The right panel of Figure 2, on the other hand, offers a compressed representation for the three words, with letters shared by the three forms activating identical *BMUs*. As a result, when the shared node 'N' is activated, one loses information of which node was activated at the previous time tick.

Figure 2: A word node tree (a) and a word node graph (b) representing German '#BEGINNEN\$', 'BEGINNT\$', 'BEGANNT\$' and '#BEGONNEN\$'. Vertices are specialized nodes and arcs stand for weighted connections. '#' and '\$' are, respectively, the start-of-word and the end-of-word symbol.

Let us briefly consider the implications of the two *BMU* structures for word processing. In the word-tree (left), *BEGINNEN*, *BEGINNT*, *BEGANNT* and *BEGONNEN* are perceived by the map as distinct forms at the earliest branching point in the hierarchy (the 'G' node). From that point onwards, the four

words activate three distinct node paths, which further bifurcate into a fourth path at the point where *BEGINNEN* and *BEGINNT* become distinct. Clearly, whenever a node has one outgoing connection only, the TSOM has no uncertainty about the ensuing step to take, and can anticipate the upcoming input symbol with certainty. In the word-graph on the right, on the other hand, branching paths converge to the same node as soon as the four input forms share an input symbol. Having more branches that converge to a common node increases the processing uncertainty by the map. The node keeps memory of many preceding contexts, and its possible continuation paths are multiplied accordingly. As we will see in the following section, this dynamic plays a key role in word acquisition and paradigm organization with TSOMs.

7.3 Processing inflection

7.3.1 The pace of acquisition

Token frequency is known to pace the acquisition of content words (i.e. excluding function words such as articles, prepositions and conjunctions), particularly at early stages of language development (Huttenlocher et al. 1991; Goodman et al. 2008; Rowe 2012). It is widely accepted that speakers have fairly accurate knowledge of the relative frequencies with which individual verbs appear in different tenses, or with different combinations of person and number features (Ellis 2002). Even if it is clear that speakers do not actively engage in consciously counting features, they nevertheless are very good at estimating frequency distributions and their central tendencies.

Early acquisition of frequent words is generally understood to be a memory effect: the more frequently a word is input, the more deeply entrenched its storage trace in the speaker's mental lexicon, and the quicker its access. In TSOMs, word processing is intimately tuned to word storage. Nodes that are repeatedly activated by an input string, become increasingly specialized for processing that string. At the same time, they are the memory nodes used for its long-term representation. The reason why frequent words are acquired at earlier stages can be better understood in terms of this processing-storage dynamic.

To investigate how word frequency distributions affect the acquisition of inflected words in a discriminative recurrent neural network, Marzi and colleagues (Marzi et al. 2014, 2016, 2018) ran a series of experiments where the inflectional systems of various languages (English, German, Italian, Modern Greek, Spanish, Standard Modern Arabic) are acquired by a TSOM trained on different frequency distributions of the same data. In the experiments, the acquisition of each input

word form is timed by the epoch when the word is recalled correctly from its memory trace on the map.⁷

As an example, Figure 3 shows the pace of acquisition of verb forms from three German paradigms: a regular one (*brauchen* ‘need’), a moderately irregular one (*wollen* ‘want’), and a highly irregular one (*sein* ‘be’). In the plots, circles represent word forms, ranked on the y-axis by increasing frequency values, and plotted on the x-axis by their epoch of acquisition. Each panel shows the outcome of two training regimes: white circles correspond to forms that are presented to a TSOM 5 times each (uniform training regime); black circles correspond to the same forms when they are presented to a TSOM with their (scaled) frequency distributions in a reference corpus (realistic training regime). Acquisition times are averaged over 5 repetitions of the same experimental condition in the two training regimes.

Figure 3: TSOMs’ learning epochs for *brauchen* ‘need’, *wollen* ‘want’, *sein* ‘be’ with uniform (white circles) and realistic (dark circles) training conditions. For each form, frequencies are given in brackets.

We observe a significant interaction between the pace of word acquisition in the two training regimes and the degree of inflectional (ir)regularity of a paradigm. In a regular paradigm like *brauchen* (Figure 3, left panel) word forms are acquired by a TSOM at about the same epoch: between epoch 13 and 15 in a

⁷ Intuitively, a word memory trace in a TSOM is the “synchronous” union set of the map activation patterns (MAPs) for all symbols making up the word, or Integrated Activation Pattern (IAP). For a sequence of symbols to be recalled accurately from a memory trace, each *BMU* must contain detailed information about each symbol and its position in the target word. If either information is wrong, the sequence is wrongly recalled.

uniform training regime (white circles), and between epoch 14 and 17 in a realistic training regime (black circles). Note that, on average, words are learned earlier when they are uniformly distributed. The latter effect is perceivably reduced in *wollen* (Figure 3, center panel), and reversed in *sein* (Figure 3, right panel), where realistically distributed items (black circles) tend to be learned more quickly than when the same items are presented five times each (white circles). Furthermore, the time span between the first and the last acquired form in the same paradigm is fairly short in *brauchen*, longer in *wollen*, and even longer in *sein*.

Prima facie, these results are compatible with a dual-route account of regular and irregular inflection (e.g. Pinker and Ullman 2002). One could argue that the short time span taken to acquire regular forms supports the nearly instantaneous application of a general rule to the paradigm, following the acquisition of its stem. Likewise, the prolonged time span taken for the acquisition of an irregular paradigm is evidence that irregular forms are memorized in a piecemeal, itemized fashion. However, since no rule learning is in place in a TSOM, a different generalization mechanism must be invoked to account for this evidence.

In a TSOM, neighboring words (e.g. *walking* and *walked*, or *walking* and *speaking*), trigger partially overlapping memory traces, i.e. integrated activation patterns that share a few nodes. Entrenchment of shared nodes benefits from cumulative exposure to redundant input patterns, making a TSOM sensitive to sublexical structures in the input. Conversely, non-shared nodes in partially overlapping memory traces compete for synchronous activation primacy in processing, and play an important role in extending inflection analogically across paradigms.

To understand how inter-paradigmatic analogical extension takes place, suppose that the (regularly) inflected form *walking* is presented to a TSOM for the first time. Its overall integrated activation pattern takes advantage of full activation of the stem *walk-* from – say – the already known *walks*, and the activation of *ing*-nodes trained on other forms like *speaking* or *making*. Note that *ing*-nodes will compete with the *s*-node of *walks*, prospectively activated by the forward temporal connections emanating from *walk*-nodes. A successful generalization ultimately depends on the final outcome of this competition, based on the synchronous activation of map nodes by the two different layers of TSOM connectivity: the input layer and the temporal layer. The comparatively faster process of acquisition of the inflected forms in a regular paradigm takes place when all inflectional endings are repeatedly seen across several paradigms,⁸ and the paradigm-specific stem is

---

⁸ We discuss processing effects of the frequency distribution of forms sharing the same inflectional ending (or inflectional entropy) in Section 7.3.2.

already acquired. This generalization step is markedly more difficult within irregular paradigms, where more stem allomorphs are available and compete with one another for activation primacy (e.g. *sings*, *sang*, *sung*). This is a gradient irregularity effect: the more stem allomorphs are found in a paradigm, the longer the time needed to acquire and associate them all with their endings.

Note finally the important difference in the pace of acquisition between the two distributions of *sein*, in sharp contrast with the strong correlation between the two distributions of *brauchen*. Despite the difference in the order of magnitude between the two distributions, the comparative insensitivity of regulars to frequency effects is reminiscent of a regularity-by-frequency interaction (Ellis and Schmidt 1998). In inflection, being more regular means being repeatedly attested in large classes of verb paradigms, and, within each paradigm, across all paradigm cells. This is not the case for irregular inflection. A radically suppletive paradigm like *sein* makes it hardly possible to infer an unattested form from other members of the same paradigm. The majority of *sein* forms are acquired one by one, as an inverse function of their own frequency distribution (and length): high-frequency items are learned earlier (Figure 3, black circles in the top left corner of the rightmost panel), and low-frequency items tend to be learned later. Conversely, in a paradigm like *brauchen*, the cumulative token frequency of the invariant stem *brauch-* compensates for the varying rate at which individual forms are shown in input. This is also true of affix allomorphs. The frequency of regular affixation is multiplicatively affected by an increase in the number of lexemes in training. A doubling in the number of verb entries in training results (if we ignore phonological adjustments) in doubling the number of affixes they select. With irregular verbs, this is hardly the case, as affix allomorphs thinly spread across irregulars, clustering in small subclasses, and incrementing their frequency rather unevenly. Thus, the pace of acquisition of regular paradigms will be less sensitive to token frequency effects of single forms, since it can benefit from the cumulative boost in frequency of other forms in the same paradigm (for stems) and other forms of other paradigms (for affixes).

All these competition effects point to the important role that paradigm entropy plays in word processing/learning. We already mentioned paradigmatic entropy in connection with child acquisition of inflection (Section 6.2).

7.3.2 Inflectional regularity: A processing-oriented notion

How degrees of inflectional regularity affect processing strategies has been recently evaluated cross-linguistically by looking at the way a TSOM predicts an upcoming input form in word recognition (Marzi et al. 2018; Marzi, Ferro and Pirrelli 2019).

The panels in Figure 4 illustrate the dynamic of word access for verb sets of 6 languages: Standard Modern Arabic, English, German, Modern Greek, Italian and Spanish, which offer evidence of graded levels of morphological (ir)regularities and complexity. For each language, they plot how easily regularly vs. irregularly inflected verb forms are predicted by a TSOM trained on 50 high-frequency sub-paradigms, each containing 15 uniformly distributed forms. Symbol prediction is graphed by the distance of each symbol to the morpheme boundary between the word stem and its inflectional ending (centered on the first symbol of the ending: $x = 0$). Each plot shows, on the y-axis, how well a symbol is predicted by the map on the basis of its preceding context. Symbol prediction in regular forms is plotted by green dashed lines, and symbol prediction in irregular forms by red solid lines.

In all tested languages, the stem prediction rate increases more steadily in regulars than irregulars (negative values on the x-axis, p-value <.001). The trend reflects the reduction in uncertainty that the map experiences when serial information is received (like in spoken word recognition). As the signal unfolds, the set of possible competing sequences that are compatible with the signal narrows down, until the point is reached when only one candidate can match the incoming signal (so-called “uniqueness point” in Marslen-Wilson’s cohort model, 1990). For sure, not all stem allomorphs of a paradigm are equally likely to be compatible with an incoming signal at any given point in time. Competition among available candidates is modulated by frequency, with more frequent allomorphs being more entrenched and so more likely to win over their competitors. The result is that the map is slower to recognize low frequency allomorphs with high-frequency competitors, as the latter are harder to eliminate (Lively, Pisoni and Goldinger 1994; Luce 1986; Luce and Pisoni 1998).

Turning back to Figure 4, another significant cross-linguistic effect is the drop in the prediction rate at the morpheme boundaries ($x = 0$), in both regularly and irregularly inflected forms. We take such a level of discontinuity at the morpheme boundary to reflect the map’s sensitivity to syntagmatic word structure. Harris’ *Mathematical Structures of language* (1968) describes how sublexical structures can be segmented out of a continuous string of symbols using the sequential likelihood between adjacent symbols. This work provided the foundations for much of later information-theoretic work on the bootstrapping of linguistic knowledge from unsupervised data (Brent 1999; Christiansen et al. 1998; Juola 1998). Our evidence is in keeping with this work. Since our training data contain no information about morphological structure, word structure emerges as an effect of specialization of processing nodes through learning. By being repeatedly exposed to input sequences, *BMUs* develop a context-sensitive representation of input symbols through incoming temporal connections, while developing at the same time strong expectations for symbols yet to come, through

Figure 4: For each language set, regression plots of interaction effects between morphological (ir)regularity and distance to morpheme boundary, in non-linear models (GAMs) fitting the number of symbols predicted by TSOMs. Categorical fixed effect is regularity (green dashed lines) vs. irregularity (red solid lines).

outgoing temporal connections. In the end, discontinuity in the strength of local temporal connections correlate with (graded) levels of sublexical structure.

The drop of prediction rates at morpheme boundaries tends to be more prominent for regular stems than for irregular ones. This is a clear paradigm effect. Stem allomorphs are found in a few paradigm cells only, and select a subset of the inflectional endings available in their paradigm. This reduces the amount of uncertainty at the morpheme boundary, as shown by the difference in prediction drop between regulars and irregulars in Figure 4. In particular, in Arabic irregular paradigms, inflectional endings are strongly predicted as a consequence of being cued by inflecting prefixes. Interestingly, from a cross-linguistic perspective, it can be observed that discontinuous patterns, typically attested in irregular paradigms of concatenative languages and systematically attested in non-concatenative morphologies such as Arabic, tend to require a higher processing cost of stems, and a lower cost in processing inflectional endings (Hahn and Bailey 2005).

These results provide evidence that perception of morphological structure crucially interacts with formal transparency and regularity in all languages. As a general trend, sublexical constituents are perceptually more salient when they remain *unchanged* across different contexts. In addition, perception of structural discontinuity increases with *the number* of different contexts where constituents are found. In regular paradigms, stems and endings combine more freely than stems and endings in irregular paradigms do. Hence regulars tend to exhibit a clearer morphological structure than irregulars, which, in turn, tend to induce a more holistic processing strategy.

As observed with stems, also inflectional endings tend to be predicted better by a *TSOM* as more symbols are input. Once more, this is mainly an effect of increasingly reduced processing uncertainty due to the narrowing down of the set of possible inflectional endings compatible with the input sequence. In the end, a uniqueness point is reached, i.e. a processing point where only one candidate inflectional ending is compatible with the unfolding signal, and the whole input form is recognized. Balling and Baayen (2008, 2012) call the point at which an inflected form can be told from all other forms of the same paradigm *Complex Uniqueness Point* (*CUP* for short). They show that late *CUPs* elicit longer processing responses by human subjects than early *CUPs* do. As a result, we expect steeper prediction slopes for endings that are uniquely identified earlier, and less steep prediction slopes for endings that eliminate their potential competitors at a later stage. What we observe by looking at our data is that, most often, the facilitative effect of regularity on stem processing is partially reversed with inflectional endings. Endings that are selected by irregular stems tend to be predicted more easily than endings of regular stems. In fact, irregular stems are more discriminative for the class of endings they can possibly select,

and this speeds up processing of ensuing endings, since they exhibit earlier *CUPs* than regulars do.

8 Measuring inflectional complexity

It makes a lot of intuitive sense to claim that some languages are inflectionally more complex than others. Everybody would agree that the English conjugation system is simpler than the German system, which is, in turn, simpler than the verb system of Modern Standard Arabic. However, when we try to motivate these deceptively trivial judgements, we are faced with a number of difficulties.

Descriptive linguists have often approached the issue through comprehensive catalogues of the morphological markers and patterns attested in a given language (Bickel and Nichols 2005; McWorther 2001; Shosted 2006). According to such approaches, the complexity of an inflectional system is assessed by enumerating the category values instantiated in the system and the range of available markers for their realization. The utility of such “enumerative” complexity or *E-complexity* (Ackerman and Malouf 2013) is however dubious on many counts.

As already shown in Section 6, researchers from diverse theoretical perspectives observe that rich inflection in fact facilitates early morphological production. In competition-based (Bates and MacWhinney 1987), as well as functional (Slobin 1982, 1985) and cue-response discriminative perspectives (Baayen et al. 2011), non-ambiguous morphological paradigms such as those of Italian conjugation are argued to provide better syntactic cues to sentence interpretation, as compared, for example, to the impoverished inflectional system of English verb agreement. Biuniqueness form-meaning relationships make inflectional markers more transparent, more compositional and in the end easier to be acquired than the one-to-many mappings of morphological forms to syntactic features that are found in English, Swedish and Dutch (Phillips 1995, 1996; Dressler 2010). Some researchers (e.g. Blom and Wijnen 2006; Crago and Allen 2001; Legate and Yang 2007) have focused on the amount of finite verbs that children receive from the adult input, to observe that the high percentage of overtly inflected forms correlates with the early production of finite forms by children. In the framework of Natural Morphology, Dressler and colleagues (Bittner et al. 2003) claimed that a richer inflection makes children more aware of morphological structure, so that they begin to develop intra-paradigmatic relations sooner than children prompted by simpler systems do (as confirmed by the quantitative results in Xanthos et al. 2011).

Another argument emphasizes that the logical problem of acquiring an inflection system consists in learning not just the full range of formal markers,

but the set of implicative relations between fully-inflected forms, which allow novel forms to be deduced from known forms (or the cell-filling problem, Ackerman and Malouf 2013; Ackerman, Blevins, Malouf 2009). To illustrate, suppose we have two hypothetical inflection 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 one 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, or vice versa. In the former system, on the other hand, exposure to one form only, no matter whether in the singular or plural, would make the speaker certain about the other form. Nonetheless, both systems present the same degree of *E-complexity*.

A number of information theoretic approaches have been proposed to model inflectional complexity in terms of either Kolmogorov complexity (Kolmogorov 1965), or Shannon entropy (Shannon 1948). The idea behind Kolmogorov complexity is to measure a dataset of inflected forms with the shortest possible grammar needed to describe them, in line with the Minimum Description Length principle (Rissanen 1989) we illustrated in connection with Goldsmith's (2001) grammar evaluation metric. The approach, however, typically (but not always, see Juola 1998) implies that a definition of morphological complexity is heavily dependent on the grammar formalism adopted (Bane 2008; Sagot and Walther 2011; Sagot 2018). To obviate this, Ackerman, Blevins and Malouf (2009), and 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 minimize 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.

More recently, Bonami and Beniami (2017) propose to generalize affix-to-affix inference to inference of intra-paradigmatic form-to-form alternation patterns, along the lines of Pirrelli and Yvon (1999), Albright (2002) and Bonami and Boyé (2014). The approach offers several advantages. It avoids the need for theoretically-loaded segmentation of inflected forms into stems and affixes in the first place. Secondly, it models implicative relations between stem allomorphs (or stem-stem predictability), thereby providing a principled way to discover so-called "principal parts", i.e. a minimal set of selected forms in a paradigm from which all other paradigm members can be deduced with

certainty (Finkel and Stump 2007, among others). Finally, it emphasizes the role of joint prediction, i.e. the use of set of forms to predict one missing form of the same paradigm, as a convenient strategy to reduce the speaker's uncertainty in the cell filling problem.

To sum up, entropic scores provide extremely valuable insights into the organization of static, synchronic paradigms. Nonetheless, measuring entropic complexity is heavily dependent on the algorithmic procedures we use for producing sets of alternation patterns, and establishing what counts as a partition of paradigm cells selecting the same stem alternant. Although speakers are very good at finding redundant patterns in paradigmatically related forms, it is not clear how their “discovery procedures” can be implemented in a typologically unbiased way. Besides, there are crucial complementary questions about how such patterns are processed and acquired that have so far been relatively neglected by the linguistic literature. We contend that inflectional complexity is an inherently multi-factorial and dynamic notion, which depends on the distributions of both stem and affix allomorphy, and on their interaction in the processing of larger syntactic constructions. Most of the quantitative metrics reviewed so far appear to focus on one or two specific factors only.

We see two principled hurdles in any attempt to identify a single, overall figure of merit for morphological complexity. First, it is exceedingly difficult, if possible at all, to integrate many distribution scores into a single overall figure, approximating a comprehensive level of inflectional complexity. Secondly, it remains to be explained how such a global score can in fact govern local inference steps, such as those taken by speakers learning an inflection system, and how it can ultimately be derived from them. We suggest that data-driven computational modelling provides a unique chance to empirically evaluate to what extent systemic complexity spontaneously emerges from acquisition of concrete examples of usage of an inflection system. Using cognitively-inspired computational models of morphology learning, we can investigate the interaction of different factors by controlling these factors within independent training regimes, and by running different instantiations of our models on each such regime. Dynamic analysis of the way the performance of our learning systems is affected across training regimes can help us understand more of factor interaction. Methodologically, this is not too far from what is done in experimental psycholinguistics, with the important qualification that computational and psycholinguistic approaches assign different epistemological roles to behavioral data and underlying neurocognitive mechanisms (see Pirrelli et al. 2020, this volume; Marzi, Ferro and Pirrelli 2019). Nonetheless, in spite of some methodological differences, computer simulations of theoretically-posed but unobservable processes offer a sound way to overcome the problem of investigating

factor interaction, and provide a window on mechanisms and representations that cannot be observed directly in human subjects.

To illustrate, let us turn back to the cross-linguistic evidence in Section 7.3.2. Figure 5 plots, for each of our sample languages, a regression model for the rate of symbol prediction in serial word processing. Languages exhibit a similar trend, though with significantly different slopes (p -values $<.001$), with Greek forms being arguably the slowest to process (less steep slope), and English forms the quickest ones (steeper slope). Our evidence is in line with LCEC (Ackerman and Malouf 2013). The overall processing ease of considerably different inflectional systems appears to oscillate within a fairly limited range of variation. In our language sample, the upper bound (low processing costs) and lower bound (high processing costs) of this range are marked by English and Modern Greek respectively. When we do not consider word length as a covariate, our space of processing ease is staked out by Spanish (upper bound) and Standard Modern Arabic (lower bound) respectively.⁹

Overall, conjugations present marginal differences in the processing overhead they require, in spite of their typological diversity. Interestingly enough, their diversity is reflected by the different processing profiles exhibited by sublexical constituents in the different languages (Figure 5). Unsurprisingly, Italian and Spanish show a very similar syntagmatic profile, with growing prediction rates for endings (whose x values are >0). Since they typically exhibit very long suffixes with the infixation of a thematic vowel, the more symbols of inflectional endings are input, the easier for a TSOM to predict the symbols to come. This is mainly the effect of an increasingly reduced processing uncertainty due to the narrowing down of the set of possible inflectional endings compatible with the input sequence.

TSOM **discriminative access** can considerably benefit from a smaller selection of inflectional endings being available at the stem-ending boundary. In Arabic imperfective forms, for example, prefixation conveys person features, thus making selection of inflectional endings highly predictable, given the stem. Conversely, Arabic stems are significantly more difficult to predict, as confirmed by the smaller coefficients for both intercept and slope (p -value $<.001$) in our generalized additive model. Unlike Arabic and English, all other languages in our test sample exhibit a wider selection of inflectional endings available at the

---

⁹ Prediction across input words is calculated by incrementally assigning each correctly predicted symbol a 1-point score, i.e. the prediction score of the preceding symbol incremented by 1. Otherwise, for unpredicted symbols the score is 0. Therefore, the longer the input word, the more likely it is to be predicted. In our set of data, Spanish verb forms tend to be longer than any others.

Figure 5: Regression plot of interaction effects between languages and distance to morpheme boundary, in a GAM fitting the number of symbols predicted by TSOMs for input words (Top panel, adapted from Marzi et al. 2018), and for stems (left panel) and inflectional endings (right panel) separately.

morpheme boundary. Nonetheless, language-specific processing effects are better understood by looking at the way inflection is formally realized in each language.

In particular, German inflection presents fairly systematic processes of stem alternation, followed by a full set of embedded endings such as *-e*, *-en*, *-end* (e.g. *beginn-e*, *beginn-en*, *beginn-end*, respectively ‘(I) begin’, ‘(we, they) begin’, ‘beginning’ (present participle)). As a result, German stems and endings are predicted according to two reversed patterns, showing, respectively, a growing and a decreasing prediction rate, as plotted in Figure 5 (left and right panels). The effect can be understood in terms of the distance between the word’s Uniqueness Point (UP, Marslen-Wilson 1984; Marslen-Wilson and Welsh 1978), i.e. the point in the input where there is only one possible lexical continuation of the currently activated node chain, and the Complex Uniqueness Point, or CUP, i.e. the point where an inflected form can be distinguished from all its paradigm companions (see §7.22). The earlier the UP, the easier for a stem to be predicted. Likewise, the earlier the CUP, the better for an ending – and hence a whole input word – to be accessed. These disambiguation points have an inhibitory effect on both word access and prediction, as confirmed by evidence on reaction times in acoustic word recognition (Balling and Baayen 2008, 2012). UP disambiguates the input stem from other onset-overlapping stems (be they paradigmatically-related or not). CUP distinguishes a specific ending from any other possible candidate, i.e. it distinguishes the input form from any other paradigmatically related form. Clearly, the fewer the endings that combine with a stem, the easier their processing.

It is useful at this stage to focus on the interaction between processing complexity and inflectional regularity. In this connection, we propose investigating inflectional complexity through a continuous, graded notion of paradigmatic (ir)regularity. For each target form, we consider its “stem-family size”, i.e. the number of paradigmatically-related forms that share, with our target, the same stem. In addition, for each paradigm, we calculate its average stem-family size, defined as the average size of the stem families belonging to the paradigm. This average score provides a quantifiable, graded notion of “paradigm regularity” that can be used in place of the traditional, dichotomous classification of inflected forms as either regular or irregular.

Figure 6 shows how the two measures affect symbol prediction by a TSOM in the serial processing of fully inflected forms (top panels), as well as their stems and inflectional endings separately (respectively, left and right panels). Note that there is a clear, facilitative effect of the family size on the processing of stems only: the greater the number of forms sharing the same stem, the easier their processing (i.e. the greater the number of predicted symbols). Conversely, when we consider full-forms and endings, there is an inhibitory effect of the family size: the more inflected forms share the same stems, the greater the processing uncertainty due to

Figure 6: GAMs predicting for our set of languages (as categorical fixed effect) the number of symbols predicted by TSOMs: fixed effects are plotted separately as paradigm regularity and stem-family size for full-forms (top panels), stems (left panels), endings (right panels). In addition to these covariates, the three GAM models include as smooth effect the corresponding length: word length for the full-form model, stem and suffix length respectively for the stem and ending models.

the larger set of possible inflectional endings compatible with the input sequence. Interestingly, a graded notion of paradigm regularity, ranging from idiosyncratic to regular paradigms, going through intermediate levels of (ir)regularity, positively correlates with our task. The upshot is that regularity favors entrenchment of stems, with an average facilitative effect on processing. Paradigm entrenchment, however, has a structural price to pay, with more complex (larger) inflection systems being more difficult to process than simpler ones.

In a functional perspective, the evidence offered here 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, where all inflected forms, both within and across paradigms, present the

earliest possible uniqueness points (UP and CUP); and (ii) a learning-driven bias for a maximally generalizable inflection system, where, for each paradigm, all forms in the paradigm can be deduced from any one of its forms, or from the smallest possible set of known forms. Clearly, a maximally contrastive system would take the least effort to process, but would require full storage of all (unpredictable) items, thus turning out to be slow to learn. A maximally generalizable system, on the other hand, would be comparatively easier to learn, but rather inefficient to process, especially when it comes to low-frequency items. What we observe is that, although languages may vary in the way they distribute processing costs across each single word due to the typological variety of the inflectional processes they resort to, if we measure the per-word processing cost as a linear function approximately interpolating prediction scores for the start-of-word and end-of-word symbols, we get values that are fairly similar.

This observation is also compatible with another clear pattern shown by the data presented in this section. In each of our sample languages, the difference between the processing cost of forms in regular paradigms and the processing cost of forms in irregular paradigms shows a structure-sensitive profile. The higher processing cost of irregular stems is partially compensated by a lower cost in processing the inflectional endings selected by irregular stems. Once more, at the level of the whole word, these structural effects make the inflectional system, from an information theoretic angle, as functional as possible to possibly contrasting processing requirements.

It should nonetheless be appreciated that the facilitative effect of fully contrastive paradigms is the result of the interaction of more factors, including complexity of the paradigm, word length and frequency distribution. Comparatively small differences have mostly been observed between languages that exhibit the same (or a comparable) number of morphosyntactic oppositions, which require the same (or a comparable) amount of syntactic contexts to be checked and interpreted. When these conditions are not controlled, acquiring an inflection system with a larger number of contrasting forms may turn out to be harder than acquiring a simpler inflection system. For example, Basque verb agreement marks an inflected verb form with affixes for subject, direct object and indirect object case. The system is agglutinative, and the number of possible distinct affix combinations for ditransitive verbs soon gets very large (up to 102 different forms in the present indicative of the auxiliary). Quantitative evidence from child language inflection shows that production of root infinitive is more frequent and prolonged in Basque than it is in a less inflectionally rich language like Spanish (Austin 2010, 2012). In fact, this may be a consequence of several concomitant factors. Basque paradigms have a much larger number of cells than Spanish paradigms have. Furthermore, the amount of syntactic context that must be processed for a

child to check case assignment on the main verb form, is considerably larger than what is needed for Spanish. Once more, it is not one factor only, but the concomitant interaction of a number of factors that may be responsible for a more complex system, and account for slower inflection acquisition.

Far from dealing with the full complexity of inflectional systems across languages, we suggested here a departure from traditional approaches based on either the stocktaking of features and their markers, or a full grammatical description of an inflection system, which all seem to require a lot of knowledge about lexical/grammatical units as well as rules/processes for their recombination/merging. Computer simulations of (discriminative) inflection learning offer a novel perspective on these issues, since they do not require that formal representations are already established. Our evidence naturally prompts the view that the overall complexity of an inflectional system is the resulting equilibrium state of a number of conflicting processing requirements and adaptive responses to task-dependent pressures (see also Marzi, Ferro and Pirrelli 2019).

9 Concluding remarks

Inflection is a fundamental area of word inquiry that lies at the interface between morphology proper, i.e. knowledge of how words are shaped and internally structured, and syntax, i.e. knowledge of what syntactic contexts make certain lexical shapes obligatory. The two dimensions are logically distinct and, in principle, independent. Nothing, in the specific way words are arranged syntagmatically, impinges on the way the same words are inflectionally marked. The great variety of formal means by which identical clusters of inflectional features are marked in morphology, both cross-linguistically and within the same language, bears witness to this autonomy, and lends itself reluctantly to being cast into combinatorial patterns of morpheme arrangement. In this chapter, we mainly focused on aspects of word realization, i.e. on knowledge of the way words are inflected, and on what factors influence speakers' acquisition of this knowledge. Here, we recap a few take-home points.

Of late, the time-honored idea that word forms are organized through paradigmatic families has proved to be extremely fruitful in accounting for important effects of lexical organization and processing. Paradigms appear to organize word forms in a network of items in complementary distribution (or, as Saussure put it, *in absentia*). This network is controlled by two functionally-motivated, interacting principles. The first principle is discriminative: formal variants must be able to mark the entire space of inflectional contrasts exhibited by a language. From this

perspective, the more dissimilar an inflected form is from its paradigm companions, the more effectively it is associated with its cluster of inflectional features (or paradigm cell). However, if paradigmatically related forms were all arbitrarily different, a speaker would be in no position to interpret or produce novel forms. The second principle counterbalances this effect. It is implicational or predictive: patterns of variation tend to be interdependent in ways that allow speakers to predict novel forms on the basis of encountered forms. This is a hallmark of regular inflection. Nevertheless, even the most suppletive or least predictable paradigms in a language typically present a few implicational patterns of formal redundancy.

A discriminative/implicational account of the paradigm dimension sheds light on the graded nature of (ir)regularity and structure in inflectional systems. Morphological irregularity is not dysfunctional, but responds to a maximally contrastive function in both word recognition and production. Since irregularly inflected forms are typically isolated, and are acquired by being committed to memory, it is only to be expected that irregularity strongly correlates with token frequency. Regularly inflected forms, on the other hand, can benefit from repeated patterns of intra-paradigmatic formal redundancy and are, therefore, also sensitive to family size (or type frequency) effects.

Any inflectional system of average complexity typically presents a whole range of gradation along this continuum. Models that postulate a dichotomous classification of inflected forms between regulars and irregulars can only account for somewhat ideal cases of particularly simple inflectional systems. Both morphological theory and computational morphology have laid considerable emphasis on graded patterns of inflectional generalization governed by lexical information. For decades, this has been one of the cornerstones of Lexicalist Morphology and has guided important computational work on inheritance lexical networks such as DATR. The idea that both general and irregular inflectional patterns can be cast into formally uniform statements ultimately blurs the distinction between rules and lexical entries. In DATR, so-called lexical rules are expressed as statements containing free variables, which are bound to constant, local values within individual, idiosyncratic lexical entries.

The idea of measuring the complexity of an inflectional system in terms of inferential uncertainty (or information entropy) represents an important recent development in paradigm-based approaches to inflection. For any given verb stem s , one can estimate how easily an unknown stem-affix combination $s-a_j$ can be predicted on the basis of an already encountered stem-affix combination $s-a_k$ for the same verb. In addition, we can also estimate how much reduction in uncertainty we get in guessing $s-a_j$, when we know more stem-affix combinations for the same paradigm. Overall, formal irregularity is not randomly scattered across paradigm cells. An inflectional system tends to reduce the amount

of uncertainty in mastering it. Besides, however complex and irregular, an inflectional system tends to be organized in such a way that less predictable forms are usually more frequent than more predictable forms.¹⁰ It is such an implicatively organized system of patterns and subpatterns that effectively addresses learnability issues, by constraining an otherwise unrestricted set of combinatorial options. Such a global functional property of inflectional systems is significantly missed by purely realizational models of inflection.

A further step in the same direction is made by moving from inheritance networks to recurrent neural networks (RNNs), where the paradigmatic organization of lexical forms is responsible for coactivation and competition of concurrently stored items in word recognition and production. The step has far reaching consequences on the way we look at word knowledge, as it shifts the research focus from what speakers know when they know inflection, to how speakers develop knowledge of inflection through input exposure. According to a learning-based perspective, redundant patterns are predominantly statistical, and even irregularities appear to be motivated by their frequency distribution in the system and the general-purpose learning strategies of the human language processor. All these issues are very different in character from the formal constraints on units, representations or rule systems proposed within theoretical and computational models. Nonetheless, they offer an insightful perspective on language architecture, and shed novel light on issues of inflectional complexity.

The most influential legacy of connectionism for models of lexical processing is probably the idea that storage and processing are not segregated in functionally independent modules of the language architecture, but are better conceived of as two interlocked dynamics of the same underlying process. In processing an input stimulus, nodes respond with a short-term activation. Due to reinforcement and competitive specialization, specific nodes are trained to respond more and more strongly to a specific class of stimuli only, forming a long-term memory trace for that class. Nodes that are repeatedly fired at short time delays by the same time series of stimuli give rise to a long-term chain of nodes specialized for processing that series. To put it in terms of Hebb's law of neural plasticity, nodes used together wire together. Once more, nodes that get repeatedly activated in a

---

¹⁰ This is in fact connected with language usage and its functional relation to language change. One can hypothesise that high-frequency items are used more frequently and are thus more prone to being phonetically reduced (e.g. Bybee 2000; Jurafsky et al. 2002; Pluymaekers et al. 2005). Alternatively, it can be argued that high frequency be a consequence of the grammaticalization of a lexical item, and its resulting light functional load (Hopper & Traugott 1993). Nevertheless, common currency in language usage can play a role in protecting high-frequency irregular forms from analogical levelling (Milizia 2015).

processing routine for a specific input word, are the same units that are used for the stored representation of that word. This relatively straightforward mechanism provides a causal link between input word frequency, degrees of entrenchment of lexical representations and processing effects.

In line with neuro-anatomical evidence (Wilson 2001; D'Esposito 2007; Ma et al. 2014), RNNs model lexical working memory as the transient activation of long-term memory structures. From this perspective, word family relations are long term memory effects, based on concurrent storage of full forms. But there is more to it than just a memory effect. The two fundamental classes of redundant inflectional patterns, stems and affixes, give rise to different, interacting word families (namely, paradigms and inflectional classes), which appear to play an important role in the way individual forms are processed and perceived by the speakers.

Focusing on paradigms first, when one member of a paradigm (say *walks*) is input to an RNN, other non-target memory chains such as *walk*, *walked* and *walk-ing* are synchronously activated, due to the shared nodes associated with the common stem. The more regular the paradigm, the fewer its stem allomorphs, and the more entrenched, on average, their corresponding memory chains. Co-activation, however, raises uncertainty at the stem boundary, where many outgoing connections project their expectations for different upcoming endings. Such a degree of uncertainty can be taken to be an information-theoretic correlate of morphological structure at the level of network connectivity. High entropy paradigms increase uncertainty at the stem boundary, and make their internal structure more salient. Conversely, low entropy families develop more holistic word chains. This explains why forms in regular paradigms are perceived more compositionally than irregular ones are. Unlike irregular paradigms, where each allomorph can appear in a subset of paradigm cells only, regular stems appear throughout their paradigms.

Competition between paradigmatically unrelated forms, on the other hand, takes place with forms sharing the same inflectional ending (e.g. all *ing*-forms). In self-organizing RNNs, the strength of the connection linking – say – *speak*- and *walk*- to *ing* is controlled by the conditional probability of *speak*- and *walk*-given *-ing*. A high entropy distribution of *ing*-forms, corresponding to a more uniform distribution, favors a more balanced allocation of weights on connections at the stem boundary. Conversely, high frequency *ing*-forms tend to proportionally strengthen their connections to *-ing* nodes, weakening the corresponding connections from their low frequency competitors.

There is a clear relationship between high levels of uncertainty at the stem boundary and word processing effects. Other factors being equal, uniformly distributed members of (high entropy) inflectional families will take equal or comparable time for processing. Conversely, an unbalanced

competition, with few family members occurring much more frequently than others, results in the RNN being slower to recognize low frequency forms with high frequency competitors, as the latter are harder to eliminate (Lively et al. 1994; Luce 1986; Luce and Pisoni 1998). This explains why higher entropy families are processed more easily (Baayen et al. 2011; Bertram et al. 2000; Moscoso et al. 2004; Kuperman et al. 2010).

In realistic input conditions, inflectional endings are not distributed uniformly (Blevins et al. 2017). Hence, for any paradigm, the most balanced distribution of its members is the one where inflected forms with high frequency endings are seen more often than inflected forms with low frequency endings. In fact, this distribution strengthens the inflected forms that appear with competitive (high frequency) endings, while weakening those that select weaker endings, and this explains the role of inflectional entropy on processing (Milin et al. 2009a, 2009b).

Complex processing dynamics offer a novel perspective on assessment of complexity issues in typologically diverse inflection systems. Nowadays, computer simulations and non-linear models of data regression offer the opportunity to visualize time-bound effects of structure complexity on word processing at a considerable level of detail. One can thus inspect the non-linear trend of the processing cost of forms in both regular and irregular paradigms, as well as differential processing patterns in typologically different languages. Although languages may considerably vary in the way processing costs are apportioned within specific inflection systems, when processing costs are measured as a linear function interpolating processing ease from the start of the word to its end, linear slopes for different inflection systems are fairly comparable, suggesting that inflection systems are, ultimately, the result of a balancing act among a number of potentially conflicting requirements and adaptive responses to task-dependent pressures.

However logically independent, the paradigmatic and syntagmatic dimensions of inflection must functionally interact during language acquisition. A distributional, discriminative approach to learning appears to conflate the issues of *how* inflection is marked, and in *what contexts* marking applies. We already mentioned evidence that children learn words in chunks (MacWhinney 1976; Wilson 2003). Upon hearing contexts where the same verb is found in different morphosyntactic contexts (as in “SHE walkS” and “THEY walk”), the child is in a position to use information of the pronominal subject and the verb suffix to discriminate third person singular contexts from non-third person singular contexts, thereby discovering the relationship between S-inflection and SHE in pre-verbal position. Paradigmatic word relations are ultimately associated with contrastive points along the syntagmatic dimension. Likewise, paradigm cells can be viewed

as grammatical abstractions over this multi-dimensional space of systematic distributional contrasts in context. Clearly, this requires that longer stretches of words be concurrently committed to memory and organized through superpositional memory chains.

There is mounting evidence that this must be true. The human brain is understood to “contain” not only morphologically simple words, but also inflected and derived forms, compounds, light verb constructions, collocations, idioms, proverbs, social routine clichés and pre-compiled routinized chunks maximizing processing opportunities (Jackendoff 2002). Recognition of idiomatic expressions and multi-word units provides strong evidence of a processing system that uses all available pieces of information as soon as possible to constrain memory search and speed up processing of the most highly expected input (Grimm et al. 2017; McCauley and Christiansen 2017; Vespignani et al. 2010). Likewise, deficits in the working memory span (e.g. in children with SLD) explain difficulties in the acquisition of inflection, especially for large embedding contexts. The formal and structural similarity between periphrastic inflection and idiomatic expressions (Booij 2010; Bonami 2015) bears witness to the functional interaction between concurrent, redundant storage of multi-word chunks and inflectional marking in language acquisition. We still know comparatively little about the way this is implemented in the brain. Nevertheless, recent empirical and experimental evidence suggests that the brain might make use of relatively unlimited long-term memory resources to compensate for the relatively limited capacity of working memory (Tremblay and Baayen 2010). By storing a number of frequently needed/used multi-word units as holistic chunks, the human processor can augment its capacity by filling working memory slots with word chunks rather than individual words. More recent neuro-functional models of working memory as a limited attentional resource distributed flexibly among all items to be maintained during processing (Ma et al. 2014) can also take advantage of pre-compiled long memory chunks. In fact, since the latter are retrieved through long-term temporal connections, working memory resources can be more efficiently used to maintain inter-chunk connections. We believe that dynamic memory models such as those suggested by Ma and colleagues, which are based on the functional integration between working memory and long-term memory resources, will shift the theoretical debate away from the traditional dichotomy between word-centred vs. syntactically-oriented accounts of inflection. Future research will likely focus on scale-free mechanisms for concurrent memorization of time-series of symbols of different length, as well as scale-dependent effects of their concurrent, hierarchical organization in the human brain.

References

- Ackerman, Farrell, James P. Blevins, & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–82. Oxford: Oxford University Press.
- Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89 (3). 429–464.
- Ackerman, Farrell & Gregory Stump. 2004. Paradigms and periphrastic expressions. In Louisa Sadler & Andrew Spencer (eds.), *Projecting morphology*, 11–54. Stanford: CSLI Publications.
- Albright, Adam. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78. 684–709.
- Altmann, Gerry T. M. & Yuki Kamide. 2007. The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language* 57 (4). 502–518.
- Amha, Azeb. 2001. *The Maale language*. Dissertation University of Leiden.
- Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge: the MIT Press.
- Austin, Jennifer. 2010. Rich inflection and the production of finite verbs in child language. *Morphology* 20 (1). 41–69.
- Austin, Jennifer. 2012. The case-agreement hierarchy in acquisition: Evidence from children learning Basque. *Lingua* 122 (3). 289–302.
- Baayen, R. Harald, Tom Dijkstra & Robert Schreuder. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language* 37 (1). 94–117.
- Baayen, R. Harald & Peter Hendrix. 2017. Two-layer networks, non-linear separation, and human learning. In Martijn Wieling, Martin Kroon, Gertjan van Noord (eds.), *From Semantics to Dialectometry. Festschrift in honor of John Nerbonne*, 13–22. London, College Publications.
- Baayen, R. Harald, Peter Hendrix & Michael Ramscar. 2013. Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. *Language and Speech* 56. 329–347.
- Baayen, R. Harald, Petar Milin, Dušica Filipović Đurđević, Peter Hendrix, Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118 (3). 438–481.
- Baayen, R. Harald, James M. McQueen, Tom Dijkstra & Robert Schreuder. 2003. Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R. Harald Baayen & Robert Schreuder (eds.), *Morphological structure in language processing*, 355–390. Berlin: De Gruyter Mouton.
- Balling, Laura W. & R. Harald Baayen. 2008. Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive processes* 23 (7–8). 1159–1190.
- Balling, Laura W. & R. Harald Baayen. 2012. Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition* 125 (1). 80–106.
- Bane, Max. 2008. Quantifying and measuring morphological complexity. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*. 69–76.
- Bates, Elizabeth & Brian MacWhinney. 1987. Competition, variation and language learning. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, 157–163, Hillsdale, NJ: Lawrence Erlbaum Associates.

- Beard, Robert. 1995. *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation*. Cambridge: SUNY Press.
- Bertram, Raymond, R. Harald Baayen & Robert Schreuder. 2000. Effects of family size for complex words. *Journal of Memory and Language* 42. 390–405.
- Berwick, Robert C. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.
- Bickel, Balthasar & Johanna Nichols. 2005. Inflectional synthesis of the verb. Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures*, 94–97. Oxford: Oxford University Press.
- Bittner, Dagmar, Wolfgang U. Dressler & Marianne Kilani-Schoch (eds.). 2003. *Development of verb inflection in first language acquisition: A cross-linguistic perspective*. Berlin/New York: Mouton de Gruyter.
- Blevins, James P. 2003. Stems and paradigms. *Language* 79 (2). 737–767.
- Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42 (3). 531–573.
- Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press.
- Blevins, James P., Petar Milin & Michael Ramscar. 2017. The Zipfian paradigm cell filling problem. Ferenc Kiefer, James P. Blevins & Huba Bartos (eds.), *Perspectives on morphological organization: data and analyses*, 139–158. Leiden: Brill.
- Blom, Elma & Frank Wijnen. 2006. Development need not be embarrassing: The demise of the root infinitive and related changes in Dutch child language. Manuscript, University of Amsterdam/Utrecht University.
- Bloomfield, Leonard. 1933. *Language*. University of Chicago Press, Chicago.
- Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25, 63–110.
- Bonami, Olivier & Sarah Beniami. 2017. Joint predictiveness in inflectional paradigms. *Word Structure* 9 (2). 156–182.
- Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes. In Florence Villoing, Sarah Leroy & Sophie David (eds.), *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux*, 17–45. Presses Universitaires de Paris Ouest.
- Booij, Geert. 1993. Against split morphology. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1993*. 27–49. Dordrecht: Kluwer.
- Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1995*. 1–16. Dordrecht: Kluwer.
- Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.
- Börjars, K., Nigel Vincent & Carol Chapman. 1997. Paradigms, periphrases, and pronominal inflection: a feature-based account. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1996*, 155–180. Dordrecht: Kluwer.
- Brent, Michael R. 1999. Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences* 3. 294–301.
- Briscoe, Josie, Dorothy V.M. Bishop & Courtenay F. Norbury. 2001. Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 42. 329–340.
- Brown, Roger. 1973. *A first language: early stages*. Cambridge, MA: Harvard University Press.
- Bruck, Maggie & Stephen J. Ceci. 1999. The suggestibility of children's memory. *Annual review of psychology* 50 (1). 419–439.
- Bybee, Joan. 1988. Morphology as lexical organization. *Theoretical morphology* 119–141.
- Bybee, Joan. 1995. Regular Morphology and the Lexicon. *Language and Cognitive Processes* 10(5). 425–455.

- Bybee, Joan. 2000. Lexicalization of sound change and alternating environments. In Michael B. Broe & Janet M. Pierrehumbert (eds.), *Laboratory Phonology V: Acquisition and the Lexicon*, 250–268. Cambridge: Cambridge University Press.
- Caballero, Gabriela & Alice C. Harris. 2010. A Working Typology of Multiple Exponence. In Ferenc Kiefer, Mária Ladányi & Péter Siptár (eds.), *Current issues in morphological theory: (ir)regularity, analogy and frequency. Selected papers from the 14th International Morphology Meeting, Budapest, May 13–16*, 163–188. John Benjamins.
- Cazden, Courtney B. 1968. The acquisition of noun and verb inflections. *Child development* 39 (2). 433–448.
- Chouinard, Michelle M. & Eve Clark. 2003. Adult reformulations of child errors as negative evidence. *Journal of Child Language* 30. 637–669.
- Christiansen, Morten H., Joseph Allen & Mark S. Seidenberg. 1998. Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes* 13. 221–268.
- Clark, Eve V. 1987. The principle of contrast: A constraint on language. In Brian MacWhinney (ed.), *Mechanisms of language acquisition*. 1–33. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Clark, Eve V. 1990. On the pragmatics of contrast. *Journal of child language* 17(2). 417–431.
- Coltheart, Max, Kathleen Rastle, Conrad Perry, Robyn Langdon & Johannes Ziegler. 2001. DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review* 108. 204–256.
- Corbett, Greville G. & Norman M. Fraser. 1993. Network Morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics* 29. 113–142.
- Crago, Martha B. & Shanley Allen. 2001. Early finiteness in Inuktitut: The role of language structure and input. *Language Acquisition* 9 (1). 59–111.
- D’Esposito, Mark. 2007. From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 362 (1481). 761–772.
- de Jong, Nivja H., Laurie B. Feldman, Robert Schreuder, Matthew Pastizzo & R. Harald Baayen. 2002. The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language* 81 (1). 555–567.
- De Pauw, Guy & Peter W. Wagacha. 2007. Bootstrapping morphological analysis of Gĩkũyũ using maximum entropy learning. *Proceedings of the 8th Annual Conference of the International Speech Communication Association (INTERSPEECH 2007)*, Antwerp. 1517–1520.
- DeLong, Katherine A., Thomas P. Urbach & Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8 (8). 1117–1121.
- Dressler, Wolfgang U. 2005. Word-formation in natural morphology. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*, 267–284. Dordrecht: Springer.
- Dressler, Wolfgang U. 2010. A typological approach to first language acquisition. In Michèle Kail & Maya Hickmann (eds.), *Language acquisition across linguistic and cognitive systems*, 109–124.
- Ellis, Nick C. 2002. Reflections on frequency effects in language processing. *Studies in second language acquisition* 24(2). 297–339.
- Ellis, Nick C., & Richard Schmidt. 1998. Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and cognitive processes* 13 (2–3). 307–336.

- Elman, Jeffrey L. 1990. Finding Structure in Time. *Cognitive Science* 14(2). 179–211.
- Evans, Roger & Gerald Gazdar. 1996. DATR: A language for lexical knowledge representation. *Computational Linguistics* 22 (2). 167–216.
- Farmer, Marion. 2000. Language and social cognition in children with specific language impairment. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 41. 627–636.
- Ferro, Marcello, Claudia Marzi & Vito Pirrelli. 2011. A self-organizing model of word storage and processing: implications for morphology learning. *Lingue e Linguaggio* X(2). 209–226.
- Finkel, Raphael & Gregory Stump. 2007. Principal parts and morphological typology. *Morphology* 17. 39–75.
- Freudenthal, Daniel, Julian Pine & Fernand Gobet. 2010. Explaining quantitative variation in the rate of Optional Infinitive errors across languages: a comparison of MOSAIC and the Variational Learning Model. *Journal of child language* 37(3). 643–669.
- Geertzen, Jeroen, James P. Blevins & Petar Milin. 2016. The informativeness of linguistic unit boundaries. *Italian Journal of Linguistics* 28(2). 1–24.
- Gillis, Steven. 2003. A case study of the early acquisition of verbs in Dutch. In Dagmar Bittner, Wolfgang U. Dressler & Marianne Kilani-Schoch (eds.), *Development of verb inflection in first language acquisition: A cross-linguistic perspective*, 171–203. Berlin/New York: Mouton de Gruyter.
- Goldsmith, John. 2001. Unsupervised learning of the morphology of a natural language. *Computational Linguistics* 27(2). 153–198.
- Goldsmith, John. 2006. An algorithm for the unsupervised learning of morphology. *Natural Language Engineering* 12. 1–19.
- Goodman, Judith C., Philip L. Dale & Ping Li. 2008. Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language* 35. 515–531.
- Grimm, Robert, Giovanni Cassani, Steven Gillis & Walter Daelemans. 2017. Facilitatory effects of multi-word units in lexical processing and word learning: a computational investigation. *frontiers in psychology* 8. <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00555/full> (accessed 10 August 2019).
- Haegeman, Liliane. 1995. Root infinitives, tense, and truncated structures. *Language Acquisition* 4. 205–255.
- Hahn, Ulrike & Todd M. Bailey. 2005. What makes words sound similar? *Cognition* 97. 227–267.
- Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4. 451–464.
- Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Hale, K. & S.J. Keyser (eds.), *The View from Building 20*. Cambridge, Mass: MIT Press.
- Hamann, Cornelia & Kim Plunkett. 1998. Subjectless sentences in child Danish. *Cognition* 69 (1). 35–72.
- Hankamer, Jorge. 1989. Morphological parsing and the lexicon. In William Marslen-Wilson (ed.), *Lexical representation and process*, 392–408. The MIT Press.
- Harm, Michael W. & Mark S. Seidenberg. 1999. Phonology, reading acquisition and dyslexia: insights from connectionist models. *Psychological Review* 106(3). 491–528.
- Harris, Alice C. 2009. Exuberant exponence in Batsbi. *Natural Language and Linguistic Theory* 27. 267–303.

- Harris, Zellig S. 1942. Morpheme alternants in linguistic analysis. *Language* 18. 109–115.
Reprinted in Joos (1957), 109–115.
- Harris, Zellig S. 1951. *Methods in Structural Linguistics*. University of Chicago Press, Chicago.
- Harris, Zellig S. 1968. *Mathematical structures of language*. New York: Wiley and Sons.
- Hockett, Charles F. 1947. Problems of morphemic analysis. *Language* 23. 321–343. Reprinted in Joos (1957), 229–242.
- Hockett, Charles F. 1967. The Yawelmani basic verb. *Language* 43. 208–222.
- Hockett, Charles F. 1987. *refurbishing our foundations: elementary linguistics from an advanced point of view*. John Benjamins, Amsterdam.
- Hopper, Paul & Elizabeth Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press [2003²].
- Huttenlocher, Janellen, Wendy Haight, Anthony Bryk, Michael Seltzer & Thomas Lyons, T. 1991. Early vocabulary growth: Relation to language input and gender. *Developmental psychology* 27(2), 236.
- Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51. 639–671.
- Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press; Oxford.
- Jordan, Michael. 1997. Serial order: a parallel distributed processing approach. *Advances in Psychology* 121. 471–495.
- Juola, Patrick. 1998. Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics* 3(5). 206–213.
- Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The role of the lemma in form variation. *Papers in laboratory phonology VII*. 4–34.
- Jurafsky, David, Alan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Joan L. Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 229–254. Amsterdam: Benjamins.
- Kail, Robert & Laurence L. Leonard. 1986. Word-finding abilities in children with specific language impairment. *Monographs of the American Speech-Language-Hearing Association* 25.
- Kemps, Rachèl J.J.K., Mirjam Ernestus, Robert Schreuder & R. Harald Baayen. 2005. Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory & Cognition* 33 (3). 430–446.
- Keuleers, Emmanuel & Walter Daelemans. 2007. Memory-based learning models of inflectional morphology: A methodological case-study. *Lingue e Linguaggio* 6(2). 151–174.
- Kibort, Anna. 2010. Towards a typology of grammatical features. In Anna Kibort & Greville G. Corbett (eds.), *Features. Perspectives on a key notion in linguistics*, 64–106. Oxford: Oxford University Press.
- Kilani-Schoch, Marianne, Ingrida Balciuniene, Katharina Korecky-Kröll, Sabine Laaha & Wolfgang U. Dressler. 2009. On the role of pragmatics in child-directed speech for the acquisition of verb morphology. *Journal of Pragmatics* 41(2). 219–239.
- Kolmogorov, Andrej N. 1965. Three approaches to the quantitative definition of information. *Problems of information transmission* 1(1). 1–7.
- Kostić, Aleksandar, Tanja Marković & Aleksandar Bauca. 2003. Inflectional morphology and word meaning: Orthogonal or co-implicative domains? In R. Harald Baayen & Robert

- Schreuder (eds.), *Morphological Structure in Language Processing*, 1–44. Mouton de Gruyter, Berlin.
- Krajewski, Grzegorz, Lieven, Elena V.M. & Anna L. Theakston. 2012. Productivity of a Polish child's inflectional noun morphology: A naturalistic study. *Morphology* 22 (1). 9–34.
- Kuperman, Victor, Raymond Bertram & R. Harald Baayen. 2010. Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language* 62 (2). 83–97.
- Legate, Julie A. & Charles Yang. 2007. Morphosyntactic learning and the development of tense. *Language Acquisition* 14 (3). 315–344.
- Leonard, Laurence. 2014. *Children with Specific Language Impairment*. The MIT Press.
- Leonard, Laurence & Patricia Deevy. 2011. Input distribution influences degree of auxiliary use by children with specific language impairment. *Cognitive Linguistics* 22. 247–273.
- Leonard, Laurence, Patricia Deevy, Patricia Fey & Shelley Bredin-Oja. 2013. Sentence comprehension in specific language impairment: A task designed to distinguish between cognitive capacity and syntactic complexity. *Journal of Speech, Language, and Hearing Research* 56. 577–589.
- Lieber, Rochelle. 1992. *Deconstructing morphology*. Chicago: Chicago University Press.
- Lively, Scott E., David B. Pisoni & Stephen D. Goldinger. 1994. Spoken word recognition: Research and theory. In Morton A. Gernsbacher (ed.), *Handbook of psycholinguistics*, 265–301. San Diego, CA: Academic Press.
- Luce, Paul A. 1986. A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics* 39. 155–158.
- Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and hearing* 19 (1). 1–36.
- Ma, Wei J., Masud Husain & Paul M. Bays. 2014. Changing concepts of working memory. *Nature Neuroscience* 17 (3). 347–356.
- MacWhinney, Brian. 1976. Hungarian research on the acquisition of morphology and syntax. *Journal of Child Language* 3. 397–410.
- MacWhinney, Brian. 1995. *The CHILDES Project: Tools for analyzing talk*. Lawrence Erlbaum, Mahwah, NJ.
- Marangolo, Paola & Costanza Papagno. 2020. Neuroscientific protocols for exploring the mental lexicon: evidence from aphasia. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word knowledge and word usage: a cross-disciplinary guide to the mental lexicon*, 124–163. De Gruyter.
- Marcus, Gary F., Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, Fei Xu & Harald Clahsen. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* 57(4, Serial No. 228).
- Marelli, Marco & Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122 (3). 485–515.
- Marslen-Wilson, William D. 1984. Function and process in spoken word recognition: A tutorial overview. In Herman Bouma & Don G. Bouwhuis (eds.), *Attention and performance X: Control of language processes*, 125–150. Hillsdale: Erlbaum.
- Marslen-Wilson, William D. 1987. Functional parallelism in spoken word-recognition. *Cognition* 25 (1–2). 71–102.
- Marslen-Wilson, William D. & Alan Welsh. 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10. 29–63.

- Marzi, Claudia, Marcello Ferro, Franco Alberto Cardillo & Vito Pirrelli. 2016. Effects of frequency and regularity in an integrative model of word storage and processing. *Italian Journal of Linguistics* 28 (1). 79–114.
- Marzi, Claudia, Marcello Ferro, Ouafae Nahli, Patrizia Belik, Stavros Bompolas & Vito Pirrelli. 2018. Evaluating Inflectional Complexity Crosslinguistically: a Processing Perspective. In *Proceedings of 11th LREC 2018*, Miyazaki, Japan. Paper 745.
- Marzi, Claudia, Marcello Ferro & Vito Pirrelli. 2014. Morphological structure through lexical parsability. *Lingue e Linguaggio* XIII(2). 263–290.
- Marzi, Claudia, Marcello Ferro & Vito Pirrelli. 2019. A processing-oriented investigation of inflectional complexity. *Frontiers in Communication* 4. 48, 1–23. <https://doi.org/10.3389/fcomm.2019.00048>
- Matthews, Peter H. 1972. *Inflectional Morphology: a theoretical study based on aspects of latin verb conjugation*. Cambridge University Press, Cambridge.
- Matthews, Peter H. 1991. *Morphology: an introduction to the theory of word-structure*. Second Edition. Cambridge University Press, Cambridge.
- McCauley, Stewart M. & Morten H. Christiansen. 2017. Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science* 9 (3). 637–652.
- McClelland, James L. & David E. Rumelhart. 1981. An interactive activation model of context effects in letter perception: Part 1. An account of Basic Findings. *Psychological Review* 88. 375–407.
- McDonough, Joyce. 2000. On the bipartite model of the Athabaskan verb. In Theodor Fernald & Paul Platero (eds.), *The Athabaskan Languages: Perspectives on a Native American Language Family*, 139–166. Oxford: Oxford University Press.
- McNamee, Paul & James Mayfield. 2007. N-gram morphemes for retrieval. *Working Notes for the CLEF 2007 Workshop*, Budapest.
- McWorther, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* (5). 125–166.
- Milin, Petar, Laurie B. Feldman, Michael Ramscar, Peter Hendrix & R. Harald Baayen. 2017. Discrimination in lexical decision. *PLoS one* 12 (2). e0171935.
- Milin, Petar, Dušica Filipović Đurđević & Fermín Moscoso del Prado Martín. 2009a. The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language* 60 (1). 50–64.
- Milin, Petar, Victor Kuperman, Aleksandar Kostić & R. Harald Baayen. 2009b. Words and paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 214–252. Oxford University Press.
- Milizia, Paolo. 2015. Patterns of syncretism and paradigm complexity: The case of old and middle Indic declension. In Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds.), *Understanding and Measuring morphological complexity*, 167–184. Oxford: Oxford University Press.
- Montgomery, James W. 2004. Sentence comprehension in children with specific language impairment: Effects of input rate and phonological working memory. *International Journal of Language & Communication Disorders* 39. 115–133.
- Montgomery, James W. & Laurence B. Leonard. 1998. Real-time inflectional processing by children with specific language impairment: Effects of phonetic substance. *Journal of Speech, Language, and Hearing Research* 41. 1432–1443.

- Moscoso del Prado Martín, Fermín, Aleksandar Kostić & R. Harald Baayen. 2004. Putting the bits together: An information theoretical perspective on morphological processing. *Cognition* 94 (1). 1–18.
- Mulder, Kimberley, Ton Dijkstra, Robert Schreuder & R. Harald Baayen. 2014. Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language* 72. 59–84.
- Noyer, Rolf. 1992. *Features, positions and affixes in autonomous morphological structure*. Ph.D. thesis, MIT.
- Paul, Hermann. 1920. *Prinzipien der Sprachgeschichte*. Max Niemayer Verlag, Tübingen.
- Perry, Conrad, Johannes C. Ziegler & Marco Zorzi. 2007. Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological review* 114(2). 273–315.
- Penke, Martina. 2012. The dual-mechanism debate. In Markus Werning, Wolfram Hinzen & Edouard Machery (eds.), *The Oxford handbook of compositionality*, 574–595. Oxford: Oxford University Press.
- Peters, Ann M. 1997. Language typology, prosody, and the acquisition of grammatical morphemes. *The crosslinguistic study of language acquisition* 5. 135–197.
- Phillips, Colin. 1995. Syntax at two: Cross-linguistic differences. In Carson Schuetze, Jennifer Ganger & Kevin Broihier (eds.), *Papers on language processing and acquisition*, 225–282. Cambridge, MA: MIT.
- Phillips, Colin. 1996. Root infinitives are finite. In Andy Stringfellow, Dalia Cahana-Amitay, Elizabeth Hughes & Andrea Zukowski (eds.), *Proceedings of the 20th annual Boston University conference on language development*, 588–599. Somerville, MA: Cascadilla Press.
- Pickering, Martin J. & Simon Garrod. 2007. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences* 11 (3). 105–110.
- Pine, Julian M., Gina Conti-Ramsden, Kate Joseph, Elena V.M. Lieven & Ludovica Serratrice. 2008. Tense over time: Testing the Agreement/Tense Omission Model as an account of the pattern of tense-marking provision in early child English. *Journal of Child Language* 35. 55–75.
- Pinker, Steven. 1984. *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, Steven. 1989. *Learnability and cognition. The acquisition of argument structure*. Cambridge, MA: MIT Press.
- Pinker, Steven & Michael T. Ullman. 2002. The past and future of the past tense. *Trends in cognitive sciences* 6 (11). 456–463.
- Pirrelli, Vito. 2000. *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa-Roma: IEPI.
- Pirrelli, Vito & Marco Battista. 2000. The paradigmatic dimension of stem allomorphy in Italian verb inflection. *Italian Journal of Linguistics* 12 (2). 307–380.
- Pirrelli, Vito & Marco Battista. 2003. Syntagmatic and paradigmatic issues in computational morphology. *Linguistica Computazionale* XVIII–XIX. 679–702.
- Pirrelli, Vito, Marcello Ferro & Claudia Marzi. 2015. Computational complexity of abstractive morphology. In Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds.), *Understanding and measuring morphological complexity*, 141–166. Oxford: Oxford University Press.
- Pirrelli, Vito, Claudia Marzi, Marcello Ferro, Franco Alberto Cardillo, Harald R. Baayen & Petar Milin. 2020. Psycho-computational modelling of the mental lexicon: a discriminative learning

- perspective. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word knowledge and word usage: a cross-disciplinary guide to the mental lexicon*, 21–80. De Gruyter.
- Pirrelli, Vito & François Yvon. 1999. The hidden dimension: a paradigmatic view of data-driven NLP. *Journal of Experimental and Theoretical Artificial Intelligence* 11. 391–408.
- Plaut, David C., James L. McClelland, Mark S. Seidenberg & Karalyn Patterson. 1996. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* 103, 56–115.
- Plunkett, Kim & Patrick Juola. 1999. A connectionist model of English past tense and plural morphology. *Cognitive Science* 23 (4), 463–490.
- Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62. 146–159.
- Popova, Gergana. 2010. Features in periphrastic constructions. In Anna Kibort & Greville G. Corbett (eds.), *Features. Perspectives on a key notion in linguistics*. Oxford: Oxford University Press, 166–184.
- Purdy, John D., Laurence B. Leonard, Christine Weber-Fox & Natalya Kaganovich. 2014. Decreased sensitivity to long-distance dependencies in children with a history of Specific Language Impairment: Electrophysiological evidence. *Journal of Speech, Language, and Hearing Research* 57 (3). 1040–1059.
- Ramscar, Michael & Melody Dye. 2011. Learning language from the input: Why innate constraints can't explain noun compounding. *Cognitive Psychology* 62 (1). 1–40.
- Ramscar, Michael & Daniel Yarlett. 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.
- Ravid, Dorit, Emmanuel Keuleers and Wolfgang U. Dressler. 2020. Word storage and computation. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word knowledge and word usage: a cross-disciplinary guide to the mental lexicon*, 582–622. De Gruyter.
- Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory* 2. 64–99.
- Rice, Mabel L., Kenneth Wexler, Janet Marquis & Scott Hershberger. 2000. Acquisition of irregular past tense by children with specific language impairment. *Journal of speech, language, and hearing research* 43(5). 1126–1144.
- Rissanen, Jorma. 1989. *Stochastic complexity in statistical inquiry*. World Scientific Series in Computer Science. Singapore: World Scientific.
- Robins, Robert H. 1959. In defence of WP. *Transactions of the Philological Society* 58, 116–144. Reprinted in *Transactions of the Philological Society* 99. 116–144.
- Rowe, Meredith. 2012. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development* 83. 1762–1774.
- Rumelhart, David & James McClelland. 1986. On learning the past tense of English verbs. In Rumelhart, D.E, McClelland J. & the PDP Research Group (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1. 216–271, The MIT Press.
- Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2000*, 71–96. Dordrecht: Kluwer.
- Sagot, Benoît. 2018. *Informatiser le lexique: Modélisation, développement et utilisation de lexiques morphologiques, syntaxiques et sémantiques*. Habilitation, Université Paris-Sorbonne.

- Sagot, Benoît, & Geraldine Walther. 2011. Non-canonical inflection: data, formalisation and complexity measures. In *International Workshop on Systems and Frameworks for Computational Morphology*. 23–45. Springer, Berlin, Heidelberg.
- Scalise, Sergio. 1984. *Generative morphology*. Dordrecht: Foris.
- Schreuder, Robert & R. Harald Baayen. 1995. Modeling morphological processing. In Laurie B. Feldman (ed.), *Morphological aspects of language processing*, 131–156. Erlbaum.
- Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27. 379–423.
- Shosted, Ryan. 2006. Correlating complexity: a typological approach. *Linguistic Typology* (10). 1–40.
- Slobin, Dan I. 1982. Universal and particular in the acquisition of language. In Eric Wanner & Lila Gleitman (eds.), *Language acquisition: The state of the art*, 128–172. Cambridge: Cambridge University Press.
- Slobin, Dan I. (ed.). 1985. *The crosslinguistic study of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Stark, Rachel & James Montgomery. 1995. Sentence processing in language-impaired children under conditions of filtering and time compression. *Applied Psycholinguistics* 16. 137–154.
- Stump, Gregory. 2001. *Inflectional morphology: a theory of paradigm structure*. Cambridge (UK): Cambridge University Press.
- Taatgen, Niels A. & John R. Anderson. 2002. Why do children learn to say “broke”? A model of learning the past tense without feedback. *Cognition* 86(2). 123–155.
- Tremblay, Antoine & R. Harald Baayen. 2010. Holistic processing of regular four-word sequences: A behavioral and erp study of the effects of structure, frequency, and probability on immediate free recall. In David Wood (ed.), *Perspectives on formulaic language: Acquisition and communication*, 151–173. Bloomsbury Publishing.
- Vespignani, Francesco, Paolo Canal, Nicola Molinaro, Sergio Fonda & Cristina Cacciari. 2010. Predictive Mechanisms in Idiom Comprehension. *Journal of Cognitive Neuroscience* 22 (8). 1682–1700.
- Wilson, Margaret. 2001. The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review*, 8 (1). 44–57.
- Wilson, Stephen. 2003. Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language* 30. 75–115.
- Wulfeck, Beverly & Elizabeth Bates. 1995. Grammatical sensitivity in children with language impairment. *Technical Report CND-9512*. San Diego: Center for Research in Language, University of California at San Diego.
- Wurzel, Wolfgang U. 1970. *Studien zur deutschen Lautstruktur*. Berlin: Akademie-Verlag.
- Xanthos, Aris, Sabine Laaha, Steven Gillis, Ursula Stephany, Ayhan Aksu-Koç, Anastasia Christofidou, Natalia Gagarina, Gordana Hrzica, F. Nihan Ketrez, F., Marianne Kilani-Schoch, Katharina Korecky-Kröll, Melita Kovačević, Klaus Laalo, Marijan Palmović, Barbara Pfeiler, Maria D. Voeikova & Wolfgang U. Dressler. 2011. On the role of morphological richness in the early development of noun and verb inflection. *First Language* 31(4), 461–479.
- Young, Robert & William Morgan. 1987. *The Navajo Language*. University of New Mexico, Albuquerque.
- Zipf, George K. 1935. *The psychobiology of language*. Houghton-Mifflin.