TÍTULO TRABAJO FIN DE MÁSTER:

MODELLING FIRST AND SECOND LANGUAGE ACQUISITION AND PROCESSING WITH TEMPORAL SELF-ORGANIZING MAPS

AUTORA: PATRIZIA BELIK

Firmado Directora del Trabajo Fin de Máster:

TUTOR: DRA. HANNA T. SKORCZYNSKA SZNAJDER

TUTORA EXTERNA: DRA. CLAUDIA MARZI
Institute for Computational Linguistics – National Research Council, Italy

Firma Autor/a del Trabajo Fin de Máster:

Dª.:

ACKNOWLEDGMENTS

First and foremost, I would like to thank my two Supervisors: Dr. Hanna Teresa Skorczynska Sznajder and Dr. Claudia Marzi, who helped me to improve this work in every way. I am very grateful for all the help, time, and effort that has been put in this dissertation.

I also wish to thank all my teachers at UPV, who guided me through the marvelous yet tangled world of linguistics.

My Alma Mater deserves a special mention. I owe thanks to all my teachers at the University of Trento and at the University of Florence for all the knowledge I received during my undergraduate studies.

I want to thank my friends and colleagues, and all those who gave me a tremendous amount of support this year and to all those who believed in me.

Finally, I would like to express my gratitude to my family. My deepest gratitude to my parents, to whom I owe everything. Thank you for the privilege of living in Europe.

TABLE OF CONTENTS

1. INTRODUCTION

According to the vast literature on word processing and access, in perception and production, an input word may concurrently activate non-target lexical neighbors that become available for further processing stages. The growing body of psycholinguistic evidence shows how **competition based on word similarity and lexical redundancy** affect speakers' anticipation of incoming stimuli, improving lexical decision and facilitating word recognition.

In the domain of bilingual performance – in terms of first language (hereafter L1) and second language (hereafter L2) interaction - the interference from one language to the other may occur with respect to both language structure and linguistic processing, and can be noticed at the phonological and the syntactical levels, as well as in lexical borrowings (Van Heuven et al., 1998).

The main goal of this thesis is to model second language acquisition and processing with Temporal Self-Organizing Maps (TSOMs, Ferro et al., 2011; Marzi et al., 2012, 2014a, 2015, 2016; Pirrelli et al., 2014, 2015) by simulating some basic cognitive processes that govern lexical processing in the mental lexicon. In particular, I pretend to bridge the gap by extending the application of computational modelling of language acquisition in monolingual and bilingual contexts to Spanish, which has not yet been treated within the given research framework.

My specific objectives are the following:

- Explore the dynamics of lexical organization by means of computational modelling in Spanish language in monolingual regime, then contrast the results to the already existing Italian and German sets.
- Explore the dynamics of interplay between the above-mentioned three languages in contexts of partial as well as perfect bilingualism.
- Provide further evidence that type/token frequency, formal redundancy and lexical neighborhood affect perception, acquisition, and processing.

Since the software for TSOM training is yet neither available for an open-source download, nor for a web-service training, the experiments have been run by Dr. Marzi at the Institute for Computational Linguistics, ComPhys Lab[1] (Italian National Research Council, Pisa, Italia). Dr. Marzi also took over the extraction of the data. I myself compiled the Spanish dataset and analyzed the results in both L1/L1 and L1/L2 regimes.

Modelling human lexical processing must take into account adaptive mechanism of storage and representation of lexical input, in consideration of the fact that the way a speaker stores lexical information reflects the way it is perceived and dynamically processed.

With no information of morpho-syntactic or semantic features, the acquisition of language-specific orthotactic constraints[2] delineate the propensity to acquire novel words, and show how acquisitional strategies are affected by the past knowledge of language and the entrenched expectations on incoming stimuli. In this perspective, a strong expectation based on L1 influences the way L2 inputs are perceived. The entrenchment can be described as a specialization of most frequent input stimuli in highly routinized clusters aimed at the optimization of our mental resources[3].

This study is based on the work of and has been completed in collaboration with my external supervisor, Claudia Marzi, who provided me with some essential information to the completion of this dissertation, i.e. both the Italian and German datasets, as well as the experimental results that I will analyze in Chapter 5.

---

[1] ComPhys Lab investigates the interplay between language-specific processes and language-aspecific cognitive functions. Its cooperation with CNR Institute of Clinical Physiology (IFC) aims at contributing to the research related to the language and communication disorders in the field of biomedical sector. For further information: www.comphyslab.eu

[2] *Orthotactic constraints* refer to the fact that certain letter sequences are perceived as more typical in a given language. The acoustic counterpart of the Orthotactic constraints are the *Phonotactic constraints*, which considers the typicality of sound patterns, rather than letters.

[3] See 2.1 for a SLA perspective of the entrenchment mechanism explained by the Neuronal Commitment feature within the MacWhinney's Competition Model (2001).

To achieve my goal, a series of TSOM-based experiments will be run, where language acquisition is modelled as the storage of "time series of symbolic units (words) as routinized patterns of short-term node activation" (Marzi & Pirrelli, 2015).

Each input form is represented by a unique time-series of symbols (orthographical representations[4] in the present work), administered one symbol at a time. Since words are permanently coded in our long-term memory as neuron activation patterns that sequentially fire, they can be represented as time series of symbols, whose receptors are time-bound to one another through associative connections. To put it simply, the incoming words are intuitively processed one symbol at a time. This kind of sequential processing is mimicked by means of computational simulations.

In this perspective, each input word form is represented by a unique time-series of symbols that are vector-coded[5] on the input layer and administered to the TSOM one at a time.

With the goal of simulating paradigmatic acquisition and perception of morphological relations between fully-inflected word forms, an identical set of verb forms have been selected for Spanish, German and Italian, and administered in different training regime conditions, namely monolingual (L1), bilingual (L1/L1) and incremental first and second language (L1+L2) regimes. Within the framework of this study, TSOM training corresponds to the initial learning period when a random frequency-arranged words are showed to the input layer of the map. In this way, groups of nodes "specialize" in regard to a

---

[4] Phonological vs. Orthographic distinction corresponds to acoustic vs. visual word recognition. In terms of input stimuli, words can be perceived in two different ways, that is, phonologically and orthographically. By way of example, computational coding of the incoming word *drive* can be administered to the input layer of the map in the following two ways: phonetic transcription /draɪv/, transformed in $d,r,a,i,v,#, as opposed to the orthographic $d,r,i,v,e#, respectively.

[5] Geometrically speaking, a vector is essentially an arrow whose dimension is specified by coordinates. In the field of neural networks, a vector corresponds to an incoming input. See Anderson (2014) for more detail.

certain input stimulus as a result of repeated pattern activations. Such mechanism roughly corresponds to the early language acquisition in children, where morphological structure gradually emerges without the need to resort to the explicit rote rehearsal techniques, and explicit rule-based learning, typically adopted by adult second language learners.

The advantage of computational simulations is offered by the possibility of (i) gaining a deep understanding of the mutual relationship between representation (memory) and processing strategies (e.g. input perception, or word production); (ii) verifying, under controlled simulations of word stimuli, that memory structures represent the way input stimuli are perceived and coherently processed; (ii) analytically studying the developmental processes that govern the acquisition of the morphological lexicon in different languages, in different language exposure conditions; (iv) monitoring the interplay between input frequency and perception of formal redundancy.

The thesis is organized as follow: in Chapter 2, I will firstly introduce some fundamental models for language acquisition, based on psycholinguistic approaches. In Chapter 3 I will review and discuss the most influential models for lexical processing, and outline some computational architectures based on artificial neural networks. Chapter 4 will include a detailed description of the experimental method and corpora adopted for computational simulations, which will be described in detail in Chapter 5. Finally, a concluding Chapter (6) will follow, with some general remarks and considerations on prospective research.

2. FIRST AND SECOND LANGUAGE ACQUISITION MODELS

There is abundant literature concerning both first language (L1) and second language (L2) acquisition. Objectives of linguistic and psycholinguistic approaches may be focused either on competence or performance. Competence refers to the abstract knowledge of language, performance relates to the actual process of language use (Bates & MacWhinney, 1989).

Although several factors — age of acquisition, social environment, and language transfer among others — may affect second language acquisition (hereafter SLA) in regards to vocabulary extension, pragmatic competence, and so forth, it is widely assumed that L1 and L2 acquisition are strongly related processes. Learning new word forms in a L2 is an extension of what we use for acquiring words in our L1. By paraphrasing MacWhinney, the strong influence and interference that L1 has onto L2 supports the position that a model of L2 learning must take into account L1 linguistic structures (MacWhinney, 2005: 49).

In what follows here, I will briefly introduce MacWhinney's Competition Model and Unified Model, as well as Bailey's work on the effects of *wordlikeness* on language acquisition.

2.1. The Competition Model (MacWhinney, 2001)

The Competition Model is a psycholinguistic theory of language acquisition and processing, which postulates that language is interpreted on the basis of linguistic cues within the input, and that language is acquired relying on competing mechanisms in an input rich linguistic environment.

Bates & MacWhinney (1989) describe their Competition Model in this way:

> The Competition Model is a framework for the crosslinguistic study of language use. It is designed to capture facts about the comprehension, production, and acquisition of language by real human beings, across a variety of qualitatively and quantitatively distinct language types. (Bates & MacWhinney, 1989: 3)

Contrary to what was assumed by Chomsky's Universal Grammar[6] (Chomsky, 1968), the Competition Model posits that first and second language acquisition relies on cognitive universals, rather than linguistic universals (MacWhinney, 2001: 1). Both theories try to determine the "universal properties of human language" (Bates & MacWhinney, 1989), although the two of them adopt sharply distinct starting points. The claim that grammar can be explained in cognitive terms means that language depends on properties of human mind, which is modelled by experience, individual differences and culture, among others. The Competition Model can be defined as an **emergentist** theory of language acquisition, assuming simple learning mechanisms, common to other cognitive processes, which support language acquisition, with word learning and grammar acquisition seen as the final result of these processes (MacWhinney, 1999).

The Competition Model is compatible with functionalism, where *function* is assimilated to the concept of *goal*. According to this viewpoint, language is viewed as a set of goal-directed activities.

Three are the pillars that determine interaction within language learning: the input, the learner, and the context. First, *Input-Driven Learning* is one of the three commitments of the Competition Model, which contrasts the nativist approach by assuming that instead of innate properties, language is learned by means of the incoming stimuli acquired through the so-called *cues*. The *strength* of the cues corresponds to the weight of the connections between units and is determined by their own reliability and availability. Cue validity can be defined as follows: "cue

---

[6] The fundamental postulate of Universal Grammar is the genetic component of the language faculty, with a set of structural rules that are innate in humans, independent of experience.

validity, that is, the information value of a given linguistic device as a cue to an underlying meaning or intention" (MacWhinney & Bates, 1989: 29).

Cue validity measures the predictability in adult language processing, as well as children language acquisition (MacWhinney & Bates, 1989: 34). The higher the predictability, the better the cue.

In addition, four other dimensions contribute to cue strength (MacWhinney, 1992):

1. *Task Frequency*. The basic principle is that the frequency of the task will determine the cue's strength and ultimately enhance the validity of the latter. For example, the determination of the agent of a transitive verb is an easier task compared to anaphoric reference determination (MacWhinney, 1992:122).

2. *Availability*. The availability expresses whether a cue is present or not. For example, in Italian, the cue suggesting a preverbal position to the agent of a transitive verb has a low availability because the omission of the subject is quite frequent. This is not the case in English.

3. *Simple Reliability*. A cue is said reliable when it leads to the correct conclusion. To get back to the previous example, the cue suggesting a preverbal position to the agent of a transitive verb is unreliable in the Italian language, whereas it is highly reliable in English.

4. *Conflict Reliability*, called Conflict validity in the previous version of this model, refers to the situation in which different cues are competing and ultimately one particular cue wins over all the other available cues for the same task. Again, this can be illustrated with the example provided by MacWhinney (1992): in Dutch, case cue of a noun phrase will dominate on its preverbal position of the noun phrase.

The goodness-of-fit criteria to find the winner is determined by all the above-mentioned variables. In principle, every cue is assessed in comparison to other available cues, which means that the final result is obtained through the competition process among the existing alternatives. That is to say, the selection

of the best possible linguistic unit within any linguistic task is accurately calculated on the bases of the above-mentioned set of criteria.

The second component of the Competition Model listed below is the Learner. It addresses the effect of individual differences on language acquisition. The human brain is better described as a neural network in which neurons "fire information" to each other through axons. This vision of brain relies on the assumption that neural activity is performed through associative networks. In brief, mental processing is viewed as a highly interactive and competitive net of units. Within this dimension of the Competition Model, at least five distinct features must be outlined (MacWhinney, 1992, 2001):

1. *Transfer*. As already mentioned, within brain structure neurons are arranged into layers of connections. Therefore, we can say that the architecture of the human brain is intrinsically interactive, allowing the interconnection of its building blocks, that is to say, neural units. Some scholars (Chomsky, 1980; Fodor, 1983) split language into separate cognitive modules. However, connectionist theories in general and the Competition Model, in particular, argue that even if "a certain limited form of emergent linguistic modularity is achieved" (MacWhinney, 1992: 120), the resulting emergent modules are far from being encapsulated. As a result, the transfer of information takes place when, let's say, a second language is being acquired. The transfer component can be resumed as follows: "[…] whatever can transfer will." (MacWhinney, 2004: 18). The general assumption in the field of second language acquisition is that L2 is parasitic on L1 in a number of ways. Accordingly, L2 relies on transfer mechanisms to build its own linguistic structure to the detriment of L1. MacWhinney (2004) lists at least four classes of transfer, as follows: Sentence Comprehension, Transfer in Pragmatics, Transfer in Morphology, and Transfer in Sentence Production. For example, in Sentence Production L2 learner will use L1 articulatory patterns to pronounce L2 sentences, which is quite logical if we think that the learner's phonetic inventory completely depends on his/her native language. Zhao & Li (2007) establish a relation between parasitism and decreasing plasticity under late learning, the latter being

likely to decrease as a separate system of L2 linguistic representations is gradually built up. Besides, the evidence mentioned in the same study shows that early L2 learning leads to a lesser degree of L2 parasitism on L1, as the plasticity of the brain allows L2 to occupy more space.

2. *Neuronal Commitment*. To say that a neural area is committed means that it has attained a defined structure by establishing a set of weights that will govern the processing. Such weights configuration points at the gradual loss of brain plasticity. In a L2 learner, such a pre-existing neural arrangement will cause "catastrophic" interference of L1 onto L2, which can only be avoided with the intercession of functional neural circuits and transfer.

3. *Automaticity*. The basic principle underlying automaticity is that highly recurrent tasks will lead to a certain level of automaticity in lexical and syntactic access. The automaticity of linguistic retrieval is apparently slower in bilinguals compared to monolinguals, when both languages are involved in the lexical decision tasks (Segalowitz & Hulstijn, 2005).

4. *Functional Circuits*. Within the framework of the Competition Model, Functional Neural Circuits account for the hindering of L2 parasitism and negative transfer on L1. Phonological loop[7] and mental imagery[8] are two of the mechanisms embedded in the working principles of the Functional Circuits. Such mechanisms boost the reinforcement of language acquisition and processing (MacWhinney, 2001: 16). In other words:

---

[7] Phonological loop is a Working Memory (WM) component that both retains phonological information and provides a rehearsal process (as theorized by Baddeley, 1986). WM is characterized by its temporary retention capacity, and the phonological loop may neutralize the decay of verbal information retained in it (Christoffels & De Groot, 2004; De Groot & Van Hell, 2005).
[8] The concept of Mental Imagery refers to the "quasi perceptual experience" of mental representations that take place without any external stimuli (Thomas, 2017).

> Theories of the neural basis of verbal memory view this storage as involving a functional neural circuit that coordinates inputs from Broca's area, lexical storage in the temporal lobe, and additional structures that support phonological memory. Unlike local lexical maps, which are neurologically stable, this functional circuit is easily disrupted and relies heavily on access to a variety of cognitive resources. (MacWhinney, 2004: 13)

5. *Perspective-taking*, which is the last point of the Competition Model from the learner's perspective, assumes that the interpretation of a clause begins from a certain starting point, that is, from a perspective that goes beyond the positional relation of the words and ultimately achieves the conceptual representation (MacWhinney, 2001: 17).

The third point in this tripartite explanation of the Competition Model is the Context. The interactional context is seen as an essential component of the language acquisition process, both in L1 as well as in L2.

2.2. The Unified Model (MacWhinney, 2004)

The Unified Model is an extended formulation of the Competition Model, which incorporates an additional subset of components accounting for multilingual acquisition, transfer promotion or inhibition among others. The main difference is that the original Competition Model did not account for learning processes, whereas the Unified Model does. Learning is intended "[…] as a resonant process that relies on storage, chunking, and support to acquire new mappings." (MacWhinney, 2004: 1).

In a simplistic way, SLA consists in applying to the L2 the acquisition patterns we assimilated while acquiring our L1. In this way, L2 "borrows" the mechanisms developed during L1 acquisition. The main distinctions between the first language and the following languages acquisition lie in the fact that L1 is learned at the same time as the infant discovers the world (MacWhinney, 2004: 2), when the brain has not yet lost its original plasticity as a result of the neuronal commitment,

and when learning is supported by an intense interaction provided by the caregivers (Snow, 1999).

While some researchers maintain two different and separate processes accounting for first and second language acquisition, MacWhinney (2004) claims that some phenomena, such as transfer, suggest otherwise. For this reason, he considers more reasonable to develop a unified model of first and second language acquisition.

In Figure 1 the seven building blocks of the Unified Model are sketched, where chunking, codes, and resonance are the newly added aspects.
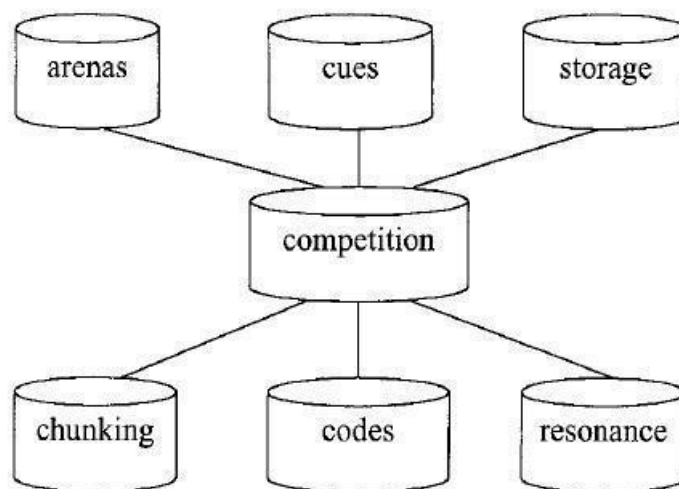
Figure 1. The Unified Competition Model general architecture (from MacWhinney, 2004).

The load-bearing component of both the earlier version of the Competition Model and the Unified Competition Model is, intuitively, competition. In the earlier model, competition was solely accounted for by the cue summation and interactive activation. Whereas, in the Unified Model competition is also based on resonance.

In summary, competitive processes take place within and between the competitive arenas, as well as between the available codes. This will be explained in some more detail in what follows.

The concept of arenas, or competitive arenas, refers to the four linguistic processing domains, namely phonology, lexicon, morphosyntax, and conceptualization. Each of them has a double representation, corresponding to the two levels of production and comprehension (e.g. phonology corresponds to the message formulation in production and to the auditory processing in comprehension). Besides, as it is to be expected, older learners develop one more competitive arena: the orthographic competition.

Cues are the other pillar of the two above-mentioned models. If we consider words as form-function mappings, cues are what allows us to establish the very connection between form and function of the linguistic sign. To put it simply, on the comprehension level, forms are cues that lead to the underlying intentions/functions; on the production level, functions are cues that lead to surface forms (forms compete to express functions).

Concerning storage, two distinct types of processing involve two distinct types of memory: offline processing mainly relies on long-term memory, with online processing mainly involving short-term memory. As emphasized by MacWhinney himself (2004), decision tasks that have no time restriction better show the validity of cue-based mechanism, since sufficient time is provided to ponder and select the best possible cue. The online processing, instead, relies on working memory and cue cost procedure supplants the cue validity mechanism. Cue cost (Bates & MacWhinney, 1987: 178) relates to the perceivability (or detectability) and assignability of the cue, the former relating to the difficulty that a listener may face in detecting cues, the latter referring to the facility to assign a role to a given cue. The assignability is then strictly interconnected with memory. In other words, cue cost increases when more processing is needed to pick up the correct cue.

One of the additional component of the Unified Model is Chunking, which operates on combinations of frequently co-occurring items to build up syllables, words, sentences. They may be found at different levels, starting from the phonology and ending with syntax. Adult learners often fail to assimilate larger inflectional patterns as a result of the tendency to pick shorter units, which are easier to analyze (MacWhinney, 2012: 15). Linguistic units are then learned on their own, instead of being learned within frequently occurring combinatorial patterns. In fact, not only chunking helps to foster fluency in language acquisition, but it can also boost the emergence of grammar by means of analogic relations.

Another additional component to the previous Competition Model is the theory of code interaction, which includes transfer theory as well. Such theory provides an explanation for how codes interact between them, that is, how the code selection happens, and what implies code switching and code mixing processes. Transfer, for its part, can be positive or negative, the former resulting from successful alignment of L2 forms with correspondent L1 forms, the latter reflecting a situation in which the alignment produces unwanted mismatches (MacWhinney, 2012: 18).

The final additional component, resonance, is central to the Unified Model in that it accounts for the code separation and learning process, among others. Structurally, resonance represents the repeated co-activation of neural connections, occurring when overt verbalizations evolve into covert, that is, inner speech that positively affects syntax reinforcement. Simply put, resonance allows the reconfiguration of oldest neural patterns. Technically speaking, resonance has to do with the co-activation of two cortical areas, which is temporarily maintained in the hippocampus (MacWhinney, 2012: 12).

In this perspective, several scenarios are outlined. For instance, in case of bilingualism, simultaneous acquisition can result into equal or unbalanced dominance of the two languages, where weaker language is unable to provide sufficient inner resonance in order to prevent transfer from happening. In contrast, balanced proficiency in the two languages inhibits the code blending. However, in comparison to bilingualism, the L2 acquisition is structurally different, because

the second language tends to be considerably weaker than the native language. In contrast to child language acquisition, in this case explicit learning strategies must be implemented.

In addition to the seven components, the Unified Model takes into account Age-Related Effects on language acquisition. There is convergent evidence that at a certain point the ability to learn languages decreases. The Unified Model mainly attributes the Critical Period Hypothesis to the entrenchment of L1, which means that during the development of L1 specific structures get determined. In such a way, the human brain commits to specific linguistic paths that increasingly lead to the limitation of brain flexibility. Self-Organizing Maps may replicate such phenomena, as will be shown in the following chapter.

Generally speaking, the Unified Model identifies two sets of factors, able of inhibiting or promoting language acquisition. Doublets of risk vs. support factors are listed as follows: entrenchment vs. resonance; misconnection vs. connection; negative vs. positive transfer; parasitism vs. internalization; isolation vs. participation.

Entrenchment and resonance duo is quite simple to understand, since the counteraction of resonance to entrenchment consists in providing a mechanism of re-encoding the already committed neuronal circuitry. Furthermore, other brain processes may increase the effect of entrenchment. The characteristic structure of the brain has the tendency to connect words that have similar meaning and grammatical category into neuron clusters, or cortical maps.

To address the connection vs. misconnection issue, one must consider the difference in part of speech assignment between L1 and L2: if both languages rely on the same system of grammatical category attribution, then no obstacle will interfere in L2 syntactic processing. On the other hand, if the two languages belong to two very different systems of grammatical category assignment, then the difficulty to learn L2 syntax will lead to a major parasitism on L1.

As already mentioned, parasitism refers to the mechanisms of emulating pre-existing L1 structures while learning L2. Internalization process has the capacity to neutralize such parasitic mechanisms with the help of the inner speech.

2.3. Generalized Neighborhood Model (GNM, Bailey & Hahn)

At this stage of inquiry, it is important to focus on the following question: how do L1 and L2 interact with each other? How does perception of typicality of lexical input affect L1 and L2 acquisition? Typicality is the extent to which a target word is perceived as similar to other words in the lexicon (Marzi et al., 2014b).

Perception of similarity is recognized to be central to language acquisition and processing as it allows to establish correspondences with other similar words in the lexicon, and, in so doing, accelerate word recognition. Following the definition of Luce (1986: 4): "A similarity neighborhood is defined as a collection of words that are phonetically similar to a given stimulus word." Similarity goes hand in hand with the concept of wordlikeness that generally refers to a speaker's knowledge of the phonotactics[9] in his/her native language, basically grounded on their lexical competence and consequently put in practice with the help of intuition (Bailey & Hahn, 2001).

Marzi et al. (2016) posit that among phenomena that affect the acquisition of verb paradigms, family effects play an important role, since the neighbor family size effect intervenes during the acquisition of low frequency regular forms. This is not the case when irregular forms are memorized, since no co-activation of neighboring connections can be performed. Both family size and the frequency of the neighboring words can positively or negatively affect the activation of the target word, e.g. if a target word is surrounded by a higher-frequency lexical units,

---

[9] Phonotactics are to be defined as the frequency of co-occurrence of certain sound patterns. Another concept to take into consideration is that phonology is governed by phonotactic constraints that identify which sound combinations are allowed and which are not in a given language (Välimaa-Blum, 2009).

it will be more difficult to access. Further evidence has shown (Baayen et al., 2006; Milin et al., 2009) that the same facilitatory/inhibitory effects are to observe in the domain of inflectional patterns, as well as inflectional classes[10], which are both involved in lexical processing.

In examining the determinants of wordlikeness, Bailey & Hahn (2001) identified two distinct factors, which oppose lexical influence, *lexical neighborhood*, to the statistical knowledge of combinatorial patterns of sequence typicality, or *phonotactic probability*. Table 1 displays Albright's (2006) review of this dichotomic model:

Lexical knowledge	Phonotactic knowledge
➢ Speakers know the words of their language ➢ Hearing a novel word activates a set of real words, while attempting lexical access ➢ The more words it activates, and the more similar it sounds to them, the more plausible it is as a possible word	➢ Speakers attend to combinatorial possibilities of different sounds in their language ➢ Novel words are parsed into constituent sounds, and the likelihood of combinations is assessed ➢ The more probable/"less illegal" the combinations are, the better the word sounds

Table 1. Generalized Neighborhood Model by Bailey & Hahn (2001)

The main main controversy underlying such distinction is whether lexical neighborhoods are fundamentally distinct from phonotactic knowledge or rather they are deeply interconnected in determining the sequence typicality. In their research paper, Bailey & Hahn (2001) set out to explore which one of the two

---

[10] Milin & Moscoso Del Prado Martín (2009) define the Inflectional Paradigm as "the set of inflected variants that can be formed for a word by regular or predictable morphological transformations." In what follows, an Inflectional Class is as well defined as "a set of words that form their IPs in the same way", which can easily be traced to what is generally called as conjugations for verbs, and declensions for nouns.

above-mentioned phenomena (or alternatively the combination thereof) determines wordlikeness, the key assumption being that there is a constant interplay between the two of them.

In the study mentioned, transition-probabilities measures are employed in order to capture the phonotactic probability at different positions across words. Transition probability is one of the two most common ways phonotactic probability is calculated, the other is bigram co-occurrence within the body of speech. Such metric allows for better exploration of the phonotactic gradience[11], which can be interpreted as a lack of definite and rigid acceptability rules (Välimaa-Blum, 2009; Bybee, 2001: 64).

On the other hand, lexical neighborhood measure is based on the phonemic overlap among lexical units within the mental lexicon, as well as the quantity of such shared phonemes (Luce, 1986). A neighbor is generally defined as "a word that could be transformed into the target word itself by a one phoneme substitution, insertion, or deletion" (Luce, 1986:17). Consequently, the neighborhood density is defined by the number of neighbor words (NNB) connected to the target word (Bailey & Hahn, 2001:571). The research taken into account here extends the measurement of neighbors to the words that feature two-phoneme edit distance.

The Generalized Neighborhood Model (GNM, from now on) is an adaptation of the Generalized Context Model (GCM; Nosofsky, 1986, 1990), which refused the neighbor vs. non-neighbor dichotomy by assuming that novel words maintain neighboring relationship of different degrees, based on the "aggregate similarity" to the stored words (Albright, 2007: 5). GNM accounts for the effects of lexical neighborhood, namely the extent to which each novel word is to some degree

---

[11] Bybee (2010) defines the concept of gradience as follows: "Gradience refers to the fact that many categories of language or grammar are difficult to distinguish, usually because change occurs over time in a gradual way, moving an element along a continuum from one category to another."

supported by the similarity to the existing words in the lexicon[12]. Such supporting effect is provided among others by token frequency, which is believed to benefit word recognition (see Luce, 1986: 5 for further discussion). Furthermore, GNM provides a more realistic measure capable of capturing both monotonic and non-monotonic frequency effects, which better determines the contribution made by words of different frequencies (Bailey & Hahn, 2001; Albright, 2007). In fact, although it seems more intuitive to assume that the frequency increase leads to stronger lexical support, it should be noted, though, that word-frequency paradox comes into play here. Word-frequency paradox (Mandler et al., 1982; Lohnas & Kahana, 2013) lies in the finding that high frequency facilitates word recalling, whereas low frequency benefits word recognition. This phenomenon is explained by Bybee (1995) as the result of the tendency of high frequency words to be processed as autonomous entities, rather than in conjunction with other lexical entries. The fact that high frequency inhibits morphological analysis results into less productivity and hence limits neighborhood interaction. This mechanism gives prominence to medium frequency words as they seem to have a greater role in determining new formations and wordlikeness (Bybee, 1995: 434; Albright, 2007: 6; Bailey & Hahn, 2001: 580).

Another newly added feature is the assessment of phonological differences, specifically aimed at identifying substitution, insertion, and deletion cost between neighbor-words. For example, in *bat-sat* the substitution of the initial consonant is easier when compared to the analogous substitution in *bat-pat* (Bailey & Hahn, 2001: 563). As a result, greater dissimilarity leads to lesser lexical support, therefore to a greater psychological distance and cost (Albright, 2007; Bailey & Hahn, 2001). Structurally speaking, the dis/similarity is calculated by means of

---

[12] Albright (2007:5) provides the following definition of the GNM:

Support for item i = $\sum$weight$_w$ × e $(-d_{i,w}/s)$, where

• weight$_w$ = a function of the token frequency of word w

• d$_{i,w}$ = psychological distance between nonce item i and existing word w

• s = sensitivity, a parameter that controls the magnitude of the advantage that very similar neighbors have in determining the outcome

• e ≈ 2.71828

For further discussion, see Albright (2007).

the natural class distance metric, which proceeds to add up all the different phonological features[13]. Such metric ranges from 0, where no difference is observed between phonemes, to 1, corresponding to completely different phonemes.

GNM's main point of departure was to determine to what extent sequence typicality is influenced by phonotactic probabilities and by lexical influences. Furthermore, the very nature of sequence typicality was questioned in terms of its submission to neighborhood effects, as opposed to the dependency on statistical knowledge (Bailey & Hahn, 2001: 585).

In conclusion, albeit acknowledging that both lexical neighborhoods and phonotactic (or orthotactic) probability are still far from being completely explored, findings from psycholinguistic and cognitive experiments run within the framework of the GNM indicate that lexical influences better predict wordlikeness than phonotactics do.

---

[13] "[…] a natural class is a group of sounds sharing one or more linguistically significant phonetic characteristic" (Bailey & Hahn, 2001: 573).

3. LEXICAL PROCESSING

Over the last forty years, many theoretical and psycholinguistic approaches have focused on the mechanism governing lexical storage and access and on the dynamics of language acquisition and processing (Jackendoff, 1975; Bybee, 1985, 1995; Matthews, 1993; Aitchison, 1994; Aronoff, 1994; Baayen et al., 1997; MacWhinney, 1999; Libben, 2005; among others).

Converging evidence from psycholinguistic studies suggests that lexical knowledge and morphological competence appear to be organized to maximize processing efficiency, rather than to minimize storage. Originally hypothesized by Vennemann (1974), the role of morphological competence is to help organize the lexicon in an appropriate way and to make words easier to be stored and accessed.

In more recent times, this hypothesis has been developed into a view of the mental lexicon as a dynamic memory system (among others, Bybee, 1995; Ellis & Schmidt, 1998; Elman, 1995, 2004; Baayen, 2007; and specifically concerning L1 and L2 interaction, see Ellis, 1998, 2002; Li, 2009).

The way lexical information is stored mirrors the way it is accessed and processed. In this perspective, an analysis of the dynamic interaction between lexical representations and distribution, degrees of regularity in input data, and perceptual competition can explain and illustrate the emergence of structure in the mental lexicon.

3.1. Mental representation and processing

According to psycholinguistic approaches, lexical representations correspond to the projection of a word in one's mental lexicon, with the latter defined as the mental lexical archive of a speaker. It is generally accepted that in such lexicon, each entry is provided with some additional information, that ranges from

linguistic features to the way that particular unit can relate to other words (Gagné, 2017; Traxler, 2012: 81)

The question to which many linguists attempted to give an answer is how words are stored, processed and retrieved in the mental lexicon. Psycholinguistic evidence has suggested different approaches to lexical representation. Gagné (2017) classifies five different approaches accounting for lexical representation:

i.   The first approach posits that no morphological parsing is involved in the first stage of word recognition: whole word forms (be they morphologically simplex or complex) are stored in the mental lexicon. The main line of argument of this model, also defined as a *full listing model* (Domínguez et al., 2000: 376), is that the lexical representation of new entries is carried out through the mapping of semantic features. However, this approach does not completely exclude morphological analysis of complex morphological units. In fact, according to some theories belonging to this approach, when accessing a new and completely unknown word, an associative mechanism activates connections between the newly input word and already stored words that are perceived as similar (Cutler, 1980), or stored words that share the same root (e.g. Granger et al., 1991).

ii. A second approach, on the contrary, relies on morphological decomposition. In this case, word representation starts with the segmentation of new lexical entries into their morphological constituents (i.e. morphemes). The general assumption of this approach, called *full parsing model*, is that the meaning of a word can be accessed by means of its isolated morphological constituents (Taft & Forster, 1976; Rastle & Davis, 2008).

iii. The third access corresponds to the *dual-route, or mixed models* (as defined by Domínguez et al., 2000: 377), which hypothesize either a parallel activation of the first two approaches described above (Schreuder

25

& Baayen, 1997), or an independent activation of the whole-word access for familiar input and the decomposing access for novel input words (Caramazza et al., 1988). In these models, a line is drawn between mainly orthographic or semantic recognition. Besides, regular and irregular forms are treated differently.

iv. A fourth approach states that the combination of the first two models operate concurrently, at the level of the surface form of the word and at the semantic level (Kuperman et al., 2009). This **interactive multiple-routes processing** claims that a chronological order must be observed, e.g. in a bottom-up processing, the surface-level must be processed before semantic level. The most important point in this approach is that the two processing routes influence each other. Recent neurophysiological evidence (e.g. Pulvermüller, 2002) suggests that surface and semantic levels get simultaneously activated when processing an input word.

v. The last approach following the overview of Gagné (2017) consists in attributing the morphological representation to the emergence of co-activation of formal features (orthographic and/or phonological patterns) together with semantic features. In other words, the representation of a given word is "dissolved", or **distributed** among a set of correlations within a network (Plaut & Gonnerman, 2000; Ferro et al., 2010; Marzi & Pirrelli, 2015).

3.2.   Morphological storage and processing models

Originally, the focus of morphological processing was placed on analytical languages, especially English, and limited to the processing model of the past tense of both regular and irregular verbs. In this work, I will revise three processing models, which account for three distinct ways to approach the morphological representation of lexical units. The main two criteria taken into

account are whether verbs are processed through storage or through computation on base-form instead.

3.2.1. DUAL PROCESSING MODEL

As already mentioned, the point of departure of morphological storage and processing models are mainly for irregular and regular verbal inflections. Pinker & Prince (1991) hold that regular morphology is directly linked to the application of rules, therefore to greater productivity, whereas irregular forms request "memory-driven processing". From such perspective, it follows that irregular versus regular patterns are subjected to two distinct and dichotomic processing mechanisms.

One assumption of this model predicts that irregular inflection is affected by the frequency rate. In fact, it claims that lower-frequency forms tend to undergo regularization (in a diachronic perspective), because their semantic representation in the mental lexicon is weaker than that of the highly-frequent verbs. Over-regularization is, in fact, a typical phenomenon not only of the U-shaped learning curve in child language acquisition (Marcus et al., 1990; Pinker & Prince, 1991: 242), but also in adult dialects, which allow coexistence of doublets such as *dived-dove* (Stemberger, 1989; Marcus et al., 1990; Pinker & Prince, 1991: 234).

Although it is accepted that the operativeness of rule-based regular inflection, consisting in online computing of inflected form by adding a suffix to the base form, does not depend on previous storage of regular forms, Pinker and Prince (1991: 237) do not exclude that the system might actually memorize some regular inflections. In fact, this may be the case when past tense doublets coexist in the mental lexicon. Again, in no case, the generalization of the regular rules depends on such storage.

3.2.2. CONNECTIONIST MODEL

Murre (2005) traces back the connectionist models to the Hebbian neural learning rule[14], often expressed as follows: ''nerve cells that fire together, wire together.'' Waskan[15] explains Hebb's proposal in the following way: "the connection between two biological neurons is strengthened (that is, the presynaptic neuron will come to have an even stronger excitatory influence) when both neurons are simultaneously active."

One of the most prominent connectionist model was theorized by Rumelhart & McClelland (1986). They suggest that:

> [...] implicit knowledge of language may be stored in connections among simple processing units organized in networks. While the behavior of such networks may be describable (at least approximately) as conforming to some system of rules, we suggest that an account of the fine structure of the phenomena of language use can best be formulated in models that make reference to the characteristics of the underlying networks. (Rumelhart & McClelland, 1987: 196)

---

[14] Hebb (1949) proposes an explanation for the adaptation of neurons in the brain during the learning process, describing a basic mechanism for synaptic plasticity, where an increase in synaptic efficacy arises from the presynaptic cell's repeated and persistent stimulation of a postsynaptic cell.
[15] Waskan, J. (2010). Connectionism. Internet Encyclopedia of Philosophy. Retrieved from: http://www.iep.utm.edu/connect/ [Last access: 11/07/2017]

According to this model, lexical representation emerges from neural connections, that is, generalization allows language acquisition. In the following citation of their work, Rumelhart & McClelland place great emphasis on one of their key assumptions:

> We have, we believe, provided a distinct alternative to the view that children learn the rules of English past-tense formation in any explicit sense. We have shown that a reasonable account of the acquisition of past tense can be provided without recourse to the notion of a "rule" as anything more than a description of the language. We have shown that, for this case, there is no induction problem. The child need not figure out what the rules are, nor even that there are rules. (Rumelhart & McClelland, 1986: 267)

The main assumption is that no distinction is to be made between regular and irregular verbs processing. The core principle of this model is that base input-forms are linked in the associative memory to the output-past-tense forms. The strengths of such connections constantly rearrange during the learning process. Such neuron-like structures, are characterized by an "all-or-none character" (Waskan, 2010), which means that they are either "firing" or they are inactive. In other words, the signals are associated with an activation value that ranges between 0 (inactive state) and 1 (maximal activation firing), they are transmitted along neural connections. The target behavior is finally achieved thanks to the adaptive learning of neural networks, constantly adjusting to the changing weight of connections. This way, both the structure and the learning process are constantly updated.

3.2.3. NETWORK MODEL

The third model of morphological processing in language acquisition that I will sketch here is the Network model, proposed by Bybee (1985; 1988). It shares some essential characteristics with connectionist models, with some fundamental new features that distinguishes it from the previous models.

On the one hand, the common ground between Bybee's theory and any other connectionist theory is that (i) ir/regular verbs are processed in the same way, contrary to what dual-processing model postulates; and, as a result, (ii) no explicit rules are formulated for language acquisition. Both models state that productivity of a morphological pattern is strictly related to its type frequency, rather than to a different morphological behavior.

On the other hand, criteria such as type and token frequency, lexical strength, and lexical schemas are some of the most relevant properties of this model, which account for distinct lexical representations and different levels of productivity of ir/regular verbs.

Lexical strength is intended as the semantic independence of a word in the mental lexicon: the stronger the representation, the less regularization tendency will emerge, the better accessibility of the entry. Bybee (1995) claims that lexical strength of a lexical item depends on its token frequency. She asserts that "words with higher lexical strength serve as the basis for the formation of new words" (Bybee, 1985).

New lexical entries create associations with other units across the lexicon, enabling the emergence of morphological connections between base forms and complex forms, as well as among complex forms. It is believed that new words do not undergo any segmentation, although a sort of morphological analysis is actually carried out in extracting common phonological and semantic features between new forms and already stored words. The number of detected common features affects lexical representation, which gradually increases as more and more morphological connections are constituted. Thus, this model posits that stronger connections result into major lexical strength. An example of how semantic and phonological connections work is sketched below (Bybee, 1988): the singular form of *cat* is semantically connected to its plural form *cats*. Furthermore, the latter form can form further connections with other plural nouns in the mental lexicon. In Figure 2, below, the relevant features are bolded for attention.
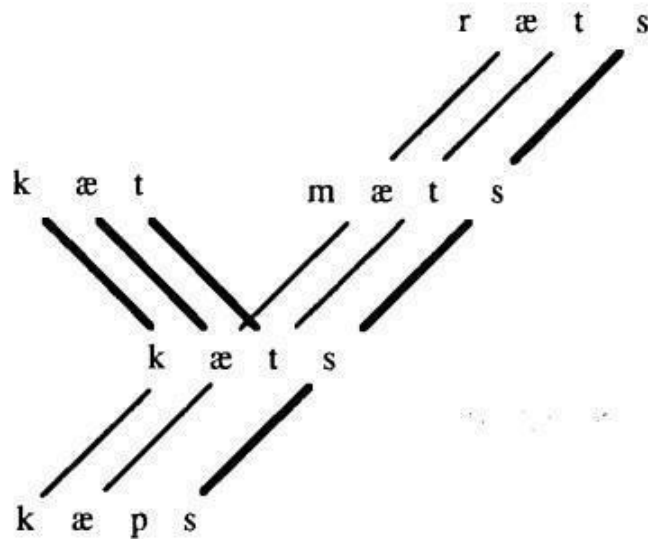
Figure 2. Lexical connections between words (Bybee, 1988: 127; 1995).

The amount of phonological and semantic connections determines the Degree of Relatedness that has a positive effect on word recognition. The author argues that the degree of relatedness depends more on shared semantic features, rather than form features (Bybee, 1988: 129).

Token frequency has a great impact on both lexical strength and lexical connections. This idea can be captured if one accepts Bybee's (1995: 129) conclusion that words with higher frequency are more autonomous, have stronger lexical representation, hence engage less with other words and are better learnt on their own terms. In contrast, less frequent forms construct a network with other lexical items, and are better learnt through such relations.

Further assumption ascribes to greater frequency the paradigm changes, such as suppletion — a common trait in the most frequent paradigms (Bybee, 1995: 129). With respect to low frequency forms, networks of shared semantic and phonological patterns result in the emergence of generalizations that Bybee

defines as *schemas*[16]. The productivity of such schemas depends on the following two criteria: (i) degree of specificity of the schema, and (ii) its lexical strength. The specificity of the schema points to its defining properties: less specific schema will be more open, thus more productive. The productivity of the schema is also positively affected by its strength, which is directly proportional to its type frequency. In this case, the type frequency refers to the occurrence of morphological pattern as a whole. The opposite scenario is as follows: the greater the number of defining features, the closer the schema. Consequently, its low productivity leads to lesser exposure to new items, and ultimately to lower type frequency.

In addition to this line of argument, Bybee (1995) makes a distinction between two sorts of schemas: product-oriented and source-oriented, which correspond to irregular and regular verbs analysis, respectively. As a matter of fact, product-oriented schemas indicate common features between base and derived forms, but do not explain how this derivation is carried out, e.g. *string-strung*. As a counterpart to this operation, source-oriented schemas do establish inflectional relation by relating the basic form with derived past from in the dental suffix, as in *walk-walked*. This latter class recalls the generative rules, theorizing online computation of past tense forms, as opposed to representation process.

Since no attempt to avoid redundancy is made according to this model, both schemas can be activated for the same connection.

3.2.4. DISCUSSION

Rule-based and classical connectionist approaches (i.e. one-route vs dual-route processing) have dominated the debate on morphological processing for a couple of decades (around from the mid-80s to the mid-2000s).

---

[16] In Bybee's approach, schemas represent phonological properties of a morphological class that are used to organize and access the lexicon.

According to the dual-route approach (§ 3.2.1), access to a morphologically complex word implicates two steps: (i) a preliminary access to the input full-word, and (ii) an optional morpheme-based access resulting from combinatorial rules.

According to the one-route approach (§ 3.2.2), morphological structure is a by-product of a direct correspondence (without combinatorial rules) between a lexical base as an input form and a correspondent output form, namely an inflected or derived form (e.g. *walk-walked, ring-rang*).

Both approaches obey to a strictly derivational view of morphological relations, according to which a fully inflected form is always produced/analyzed on the basis of a unique, underlying lexical (base) form. By modelling inflection as a phonological/orthographical mapping function from a lexical base to its range of inflected forms, classical connectionist architectures are closer to a rule-free variant of the classical constructive view (where roots and affixes are the basic building blocks of morphological competence, on the assumption that the lexicon is largely redundancy-free), than to associative models of the mental lexicon.

Nowadays, there is convergent evidence (coming from the large body of psycholinguistic studies) of a distributed account of morphological structure as an emergent property of lexical self-organization, based on relations between surface forms, in line with the network model proposed by Bybee (§ 3.2.3).

 It should be assumed that all word forms are memorized in the lexicon, with no distinction between a stored base form and all other related forms, which are processed on-line (see Baayen, 2007, for an overview).

In addition, to capture the fact that words encountered frequently exhibit different lexical properties from words encountered infrequently, any model of the mental lexicon must assume that accessing a word in some way affects the access representation of that word (Marslen-Wilson, 1993; Sandra, 1994; among others).

3.3. Computational modelling

The way lexical information is stored may reflect the way it is represented, accessed and retrieved as patterns of concurrent activation of memory areas.

Memory, as already pointed out, plays a central role in lexical modelling, and in such account, computer simulations of memory processes can well address issues of lexical acquisition and processing.

From a computational standpoint, lexical processing has to address three fundamental issues: (i) the nature of input representations, (ii) the nature of output representations, (iii) the formal relationship holding between (i) and (ii). Such issues are strongly influenced both by the way specific tasks are modelled, and by the theoretical approach they are related to, as summarized by Ellis:

> […] language researchers take recourse to computer modeling by which the test of the simulation is whether competences emerge that parallel those of human language learners exposed to similar input. In this way, the debate between deductive and inductive approaches to language acquisition is being rephrased in terms of well-articulated models and real-world data. (Ellis, 2005: 7)

Computational simulations may offer the possibility to gain insights into the relationship between representation (memory) and processing strategies (perception and production), and empirically verify, under controlled simulations of word stimuli, that memory structures represent the way external stimuli are perceived and processed in our brain.

In such account, **neural networks**, as adopted in computational modelling, represent simplified models of neural processing in the brain, as they can mimic the behavior of aggregations of neurons in the cortical areas involved in the classification of sensory data.

Kohonen's Self-Organizing Maps (SOMs, Kohonen, 2001) define a class of unsupervised[17] artificial neural networks that mimics the behavior of small aggregations of neurons in the cortical areas involved in the classification of input data.

3.3.1. KOHONEN'S SELF-ORGANIZING MAPS

In such architecture, processing consists in the activation of specific neurons upon presentation of a particular stimulus. A distinguishing feature of this brain maps is their topological organization, where nearby neurons in the map are activated by similar stimuli. There is evidence that at least some aspects of their neural connectivity emerge through self-organization as a function of cumulated sensory experience (Kaas et al., 1983). Functionally, these kind of brain maps are dynamic memory stores, which are directly involved in input processing, exhibiting effects of dedicated long-term topological organization.

Kohonen (2001) points out that artificial neural networks cannot emulate the complex anatomy of human brain, therefore they are meant to describe a limited area of it, which corresponds to a particular function. He suggested that (1982: 64) "This particular network model is first used to demonstrate, in an ultimately simplified configuration, that the activity of neighboring cells, due to the lateral interactions, can become clustered in small groups [...]".

A SOM can be defined as a "topology-preserving" map (Rojas, 1996), in so far as it succeeds in duplicating the topological correspondence, that is to say the spatial relation, between the incoming stimuli and the inputs in the cortex. For

---

[17] Unsupervised is referred to the fact that there is no human supervision, thus recreating a competitive learning process (Holmén, 1996). Such characteristic of SOMs (and TSOMs as well) reflects the lack of instructions, balanced by the ability of the map to "classify" input stimuli. In other words, perception is what enables the acquisition, where the incoming stimuli are classified, stored, and grouped in order to let shared features to emerge. In this perspective, the classification process optimizes input storage. This characteristic mimics a genuine learning process where a child is exposed to the flow of unknown linguistic information that s/he is supposed to learn without any explicit instruction. The learning process is thus carried out intuitively by detecting similarities across the incoming stimuli.

example, neurobiological evidence has shown that the human brain processes visual information in such a way that a two-dimensional mapping is projected on the cortex (Rojas, 1996: 392). The same experimental evidence points to the fact that external inputs activate the immediate neighbors of the unit.

When presenting the SOMs, Kohonen (1982: 59) explicitly referred to an "idealized neural structure", although he assumed that this model could have much wider application (Kohonen, 1982, 2001).

This kind of neural modelling realistically explains the organization of certain brain structures, whose characteristics are unfolded through SOM's adaptive processing (Kohonen, 2001: 85). In simple terms, Self-Organizing Maps originate "elastic net" of points that can be described as "ordered maps of various sensory features onto a layered neural network" (Kohonen, 2001: 69).

The nearest node to a sensory stimulus is the so-called Best Matching Unit (BMU), which is the salient unit. The competition underlying SOMs mechanism yields a winner that represents all the grouping of input vectors (Waskan, 2010). All nodes surrounding the BMU form the *neighborhood*. To sum up, during the learning process, the neighboring nodes interact in order to learn from the same sensory input. As a consequence, local relaxation on the weight of neuronal vectors within the neighborhood is achieved, which finally results in the global ordering of the map (Kohonen, 2001: 87).

Neural network modelling has been mainly applied to the domain of monolingual lexical acquisition, whereas, up to day, only a few studies implying neural modelling devoted to bilingualism has been carried out (Zhao & Li, 2007). In the attempt to bridge this gap, Zhao & Li (2007) in their study *Bilingual Lexical Representation in a Self-Organizing Neural Network Model*, focused on investigating whether within the framework of bilingualism, the lexicon of the two languages develops into one shared representation in the mental lexical, or rather they are stored separately. Zhao & Li (2007) proposed a SOM-based architecture with three feature maps, corresponding to input level (auditory-phonology),

lexical-semantic representation level, and output level (articulation). Once it receives semantic or phonological input vector, a SOM generates representational mappings on two-dimensional array of nodes that can take a rectangular or hexagonal form.
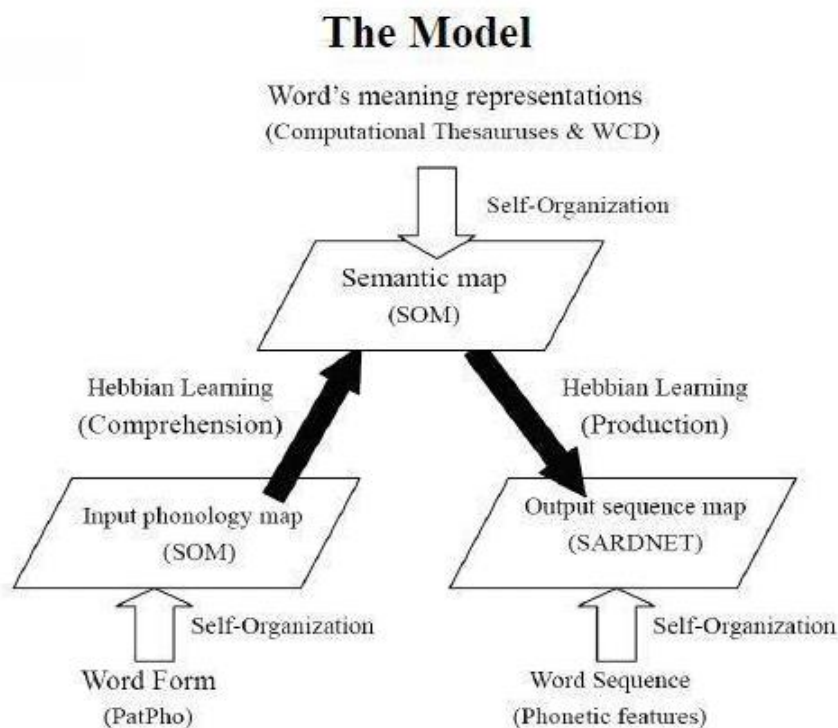
Figure 3: DevLex II structure (Li, Zhao & MacWhinney, 2007).

A whole set of variables affects language acquisition, for this reason, the above-mentioned study narrowed down the investigation by focusing on three different L2 acquisition scenarios: simultaneous, early, and late learning. The main hypothesis was that bilingual lexicon evolves progressively as L2 is being learned. For this purpose, the DevLex II, has been used in order to detect the dynamical nature of bilingual lexical organization. The structure of DevLex II is illustrated in Figure 3 above.

The three levels account for phonological, semantic, and production representations. Both the input and semantic information are represented by

SOMs, connected through Hebbian learning.

As already mentioned, DevLex II was trained to simulate three distinct learning contexts. To do so, 500 were presented simultaneously, as well as more or less sequentially, to the network over 10 learning epochs. Each word was input 10 times per epoch. The bilingualism regime was recreated by parallel exposition of the map to both lexica simultaneously, 50 form per epoch. In the second scenario, corresponding to the early learning, L2 lexicon (Chinese) was input after 100 L1 words (English) were presented to the map. Finally, within the late learning context L2 lexicon intervened after the network saw 400 L1 forms.

Figure 4 summarizes the results of such experiments by showing both phonological as well as semantic maps, showcasing the lexical distribution of L1/L2 lexica in three training regimes. Dark areas represent L2 lexicon.

In case of simultaneous learning context, the results reflect a situation in which L1 and L2 achieve two separate lexical representations, where no dominant language is attested. The boundaries of both languages are clear. On the other hand, in early learning situation L2 partially loses its agglomerate characteristic —at least in the phonology map— by giving more space to L1. In this way, L1 starts gaining dominance over L2, which can be observed also in the dedicated semantic map. In the last scenario the distributional pattern drastically changes. In fact, L2 lexicon is assimilated to the dominant L1 lexicon, L2 fragments moving closer to the areas of L1 on the basis of shared features, be they semantic or phonological. Zhao & Li (2007) explain such distribution with plasticity loss[18]. In other words, in late learning regime lexical organizational structure is already established, thus the difficulty of changing the topology of the network increases with neuronal entrenchment.

---

[18] See § 2.1 for more detail.
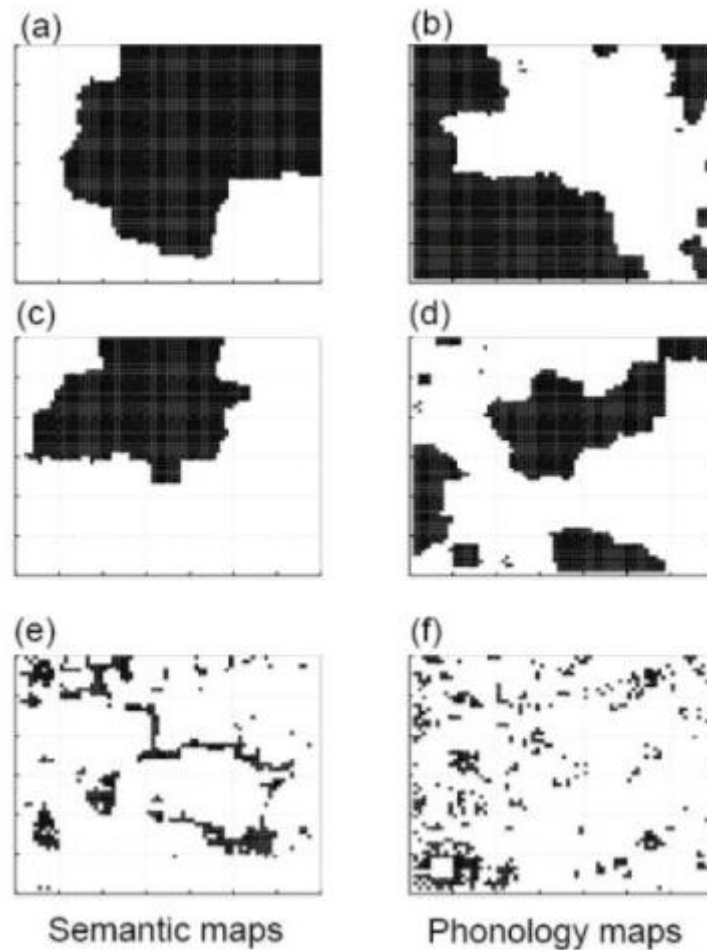
Figure 4: Bilingual lexical representations on semantic map (left) and phonological map (right). White areas correspond to L1 words (English). Black areas correspond to L2 words (Chinese), captured in the following three acquisitional regimes: simultaneous (a, b), early (c, d), and late (e, f) ones (Zhao & Li, 2007).

One limitation of DevLex II model is pertinent here: the lack of the temporal dimension of input removes the surface word relations that trigger the emergence of lexical structure. In this way, semantic representation gains at the expense of the predictability of the incoming word.

In the following section, I will briefly outline an architecture based on one level of representation, where input words are encoded on the input layer as temporal sequences of symbols.

3.3.2. TEMPORAL SELF-ORGANIZING MAPS

The main difference between classical SOMs and TSOMs is the temporal dimension: in SOMs input words are shown in a static way, whereas temporality is an essential component of Temporal Self-Organizing Maps (TSOMs hereafter). In fact, sequential representation of words as time-series of symbols is a more realistic way of recreating human language perception, be it based on acoustic, or written input.

To be more specific, TSOMs represent a variant of classical SOMs, by adding a level of temporal connectivity, implemented as a pool of re-entrant connections providing the state of activation of the map at the immediately preceding time tick. Temporal connections encode the map's expectations of upcoming input on the basis of past experience (Pirrelli et al., 2015). In this way, a map can memorize input words as time-series of symbols (e.g. orthographic letters or phonological representations) as activation chains of nodes.

The TSOM architecture (see Figure 5 below) is characterized by a grid of topologically organized memory node (exactly as Kohonen's SOMs are), which represents one layer of (artificial) neurons, with two layers of connectivity: (i) all nodes are fully connected with the input vector with no time delay (i.e. the spatial connection layer); (ii) each node is connected with all other nodes (i.e. temporal connection layer), whose connections are updated with one-step time delay, based on activity synchronization between a BMU at a preceding time and the following activating BMU (i.e. the most highly activated node at time *t-1* and the node at time *t* that mostly get activated).

Each learning step includes three phases: (i) input encoding, (ii) activation and (iii) weight adjustment. A symbol is represented on the input layer at time *t* through an input vector of *x* codes (see Figure 5). At each exposure, map nodes are activated in parallel depending on how close their weights on the spatial connection are to the *x* input vector, and how strongly nodes are connected with the BMU at time *t*-1 over the temporal layer (Ferro et al., 2011).
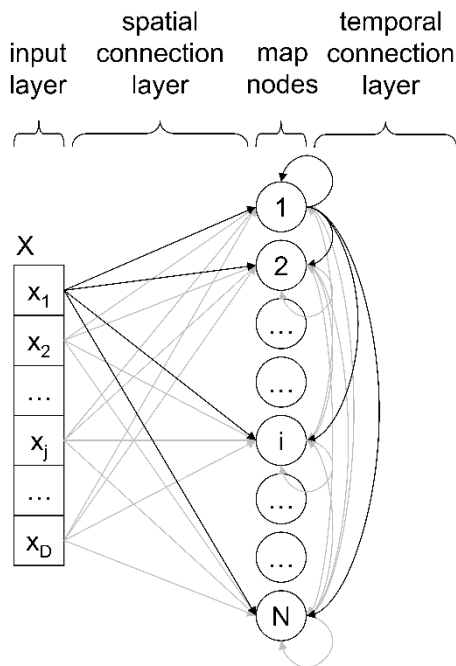
Figure 5: Outline architecture of TSOM (Ferro et al., 2011; Pirrelli et al., 2015)

During training, due to such learning dynamic, each node develops a sensitivity to both a position-specific symbol and a context-specific symbol by incrementally adjusting its weights to recurrent patterns of morphological structure. As a consequence, a pool of nodes tends to specialize to respond to any specific input symbol, by showing higher activity levels than all others when the symbols appear in a particular context (Marzi et al., 2017). In this way, the overall organization of a TSOM will be determined by the morphological structure of training data, depending on three factors: similarity, frequency, and symbol timing. In fact, similar symbol sequences generate overlapping activation patterns; highly frequent symbols and chunks tend to select dedicated nodes; nodes react differently depending on the context where a symbol is repeatedly found.

Perception of similarity between input words can thus be measured in terms of recurrent patterns shared by inflected forms, and TSOMs may provide an explanatory framework by bringing together insights from neighbor family effects

on word recognition, evidence from family size effects in serial lexical access[19] and paradigm-based dynamics in lexical acquisition.

3.3.3. MODELLING FIRST AND SECOND LANGUAGE ACQUISITION AND PROCESSING WITH TSOMS

Wordlikeness effects appear to interact with memory issues, and in particular with how lexical representations are encoded in the long-term storage. TSOMs, namely computational models of serial memories may take into account dynamics of lexical acquisition and processing, by relying on some basic mechanisms of co-activation and competition between concurrently stored words.

To make things very simple, lexical processing is based on (i) the co-activation and competition of memory resources, (ii) the activation primacy based on the "winner takes all" principle[20], and (iii) selective specialization, over training.

During acquisition, words get permanently coded in our long-term memory as neuron activation patterns that sequentially fire. In such an account, they can be conceptualized as time series of symbols.

In the architecture of TSOMs, words are represented by time-series of symbols, which are vector-coded on the input layer and administered to the TSOM one at a time. During training, words are shown randomly, without any semantic or morpho-syntactic features or any possible inter-word relations. In this way, each node gets attuned to context-specific symbol by incrementally adjusting its synaptic weights to recurrent patterns of morphological structure.

---

[19] Serial lexical access refers to the temporal dimension of lexical access, where each word is accessed through sequence of symbols. At this stage of analysis, words shall be thought of as temporal series of inputs, rather than being represented by one node only. In this way, it is possible to appreciate that also at a superficial access of row forms, namely in absence of the semantic dimension, it is possible to find out what words have in common. For example, *book* and *handbook* share part of their form, as well as *machen* and *gemacht* (German "do" and "done"), which share the stem "mach" even if it appears temporally dis-aligned in these forms, or *abriendo*, *conociendo*, etc. sharing the same inflectional suffix "*iendo*".
[20] The principle which is behind the idea of Best Matching Unit, as well.

To sum up, both short-term and long-term memory dynamics cooperate in word processing by TSOMs by triggering the short-term activation and the long-term adjustment. This two-fold dynamic[21] shows how processing and storage are mutually interdependent: node patterns get activated as input stimuli are sequentially fired, then the connections undergo readjustment that finally leads to the identification of the Best Matching Unit. At the end, the nodes that are directly connected to the BMU gain strength, those that did not get weaker. Finally, the repeated firing of the same input stimulus leads to a stronger activation of the BMU and, as a final result, to the specialization of the pattern. Within the framework of the TSOMs, the specialization of the BMU to a specific activation pattern is given by frequency distribution and formal redundancy in the training data (Marzi & Pirrelli, 2015).

The above described selective specialization simulates the human propensity to show more sensitivity towards most typical chunks in their native language. It is important to observe that during learning, connection adjustment decreases by the decreasing of plasticity. As a consequence, L2 acquisition and processing tend to be affected by a reduced specialization for context-specific symbol identity, and a weaker entrenchment of highly frequent sub-strings.

---

[21] Marzi & Pirrelli (2015) for further detail.

4. METHOD AND CORPORA

4.1. Selection criterion from corpora

4.1.1. THE SPANISH DATASET

To investigate the dynamics of word and paradigm acquisition in a bilingual training condition, a set of fifty Spanish verbs is compared with fifty German and fifty Italian verb sub-paradigms, which were selected with the same criterion[22] (respectively from the CELEX Lexical database[23], Baayen et al. 1995, and Paisà Corpus, Lyding et al. 2014).

Thus, fifty Spanish verb sub-paradigms were selected among the most highly ranked paradigms by cumulative frequency in a European Spanish subcorpus of the larger *Spanish TenTen corpus* available in the online corpus analysis tool, Sketch Engine (sketchengine.co.uk). The choice of Peninsular Spanish is motivated by the different use of the past tense between European and Latin American Spanish, which would affect the frequency of selected paradigms.

For each paradigm, an identical set of 15 cells was used for training (namely, the infinitive, present and past participle, singular and plural simple present, singular and plural simple past), for an overall number of 750 inflected forms. For each form, a function of real word frequency distributions in the reference corpus is considered (with token frequencies in the range of 1 to 1001). The correlation between corpus frequencies and dataset frequencies is significantly[24] high (r=0.99). See detailed figures in Table 2.

---

[22] The German and Italian sets of data had been selected and annotated by the *ComphysLab* at the Institute for Computational Linguistics, National Research Council of Italy.

[23] Celex is a web interface to lexical database (http://celex.mpi.nl/). For different languages (English, Duthc, German) it may contain token frequencies, part of speech tagging (POS, namely morpho-syntactic information), lemma, among others.

[24] Statistical analyses are run with R (http://cran.r-project.org). This kind of statistical analyses were provided by the Institute for Computational Linguistics (Pisa). Since it was necessary to reduce the whole amount of token frequencies for reducing the computational effort and duration in the arbitrary range of 1-1001, I found it necessary to demonstrate that the adopted function to

In case of any unattested verb form in the reference corpus, the set of the selected 15 cells was completed by adding the missing form as *hapax*.

SPANISH SET	Corpus frequencies	Dataset frequencies
Mean (standard deviation)	117681.6 (357262.1)	23.47 (66.43)
Minimum	0	1
Maximum	5376382	1001
1st quantile	1517	2
3rd quantile	95174	19
Median	24030	6

Table 2: Detailed figures for the selected dataset with reference corpus frequencies (left column) and the functionally adjusted frequencies (right column)

As expected, in both sets, word token frequencies are not normally distributed. See boxplot distribution and linear regression[25] in Figure 6.

---
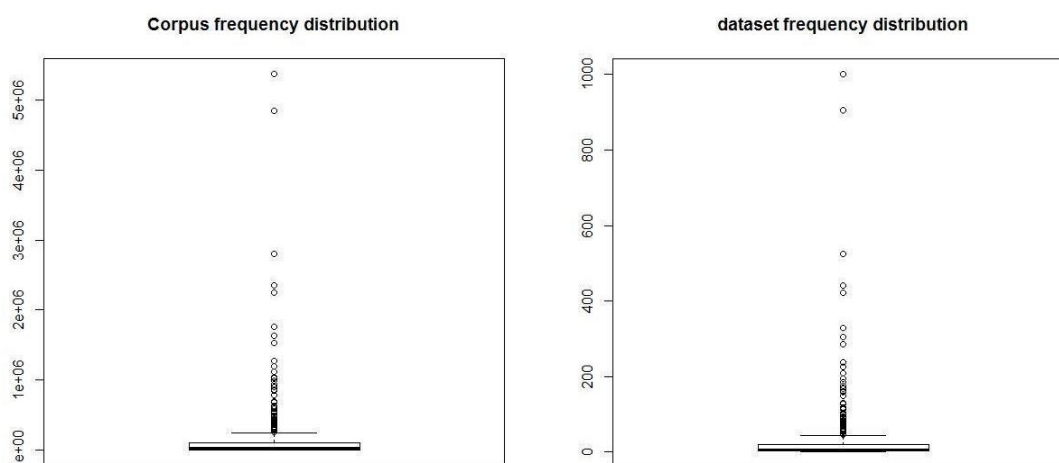
limit token frequencies did not alter the distributions of tokens, that is to say that low-frequency words are still strongly less-attested than highly-frequent ones.

[25] Linear regression is an approach for modelling the relationship between a variable (y axes, token frequency in this case) and the data (x axes). Boxplots represent a simple and intuitive way to describe data. White circles indicate outliers.
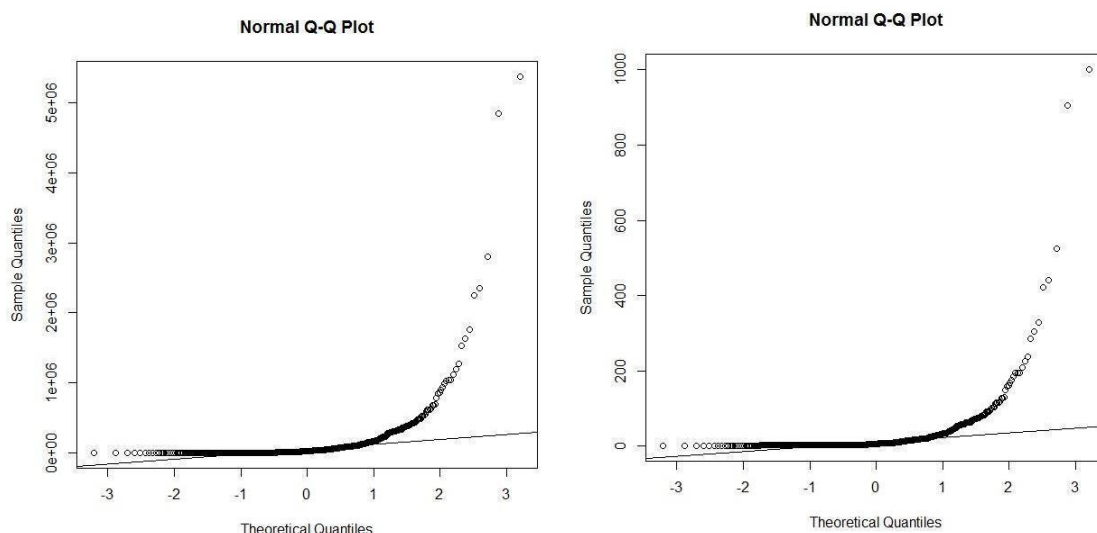
Figure 6: Boxplot distribution of Corpus (top left panel) and dataset frequencies (top right panel). Linear regressions for frequencies in the two sets, which are corpus-based (bottom left panel) and functionally reduced (bottom right panel) show non-normal distributions of token frequencies (y axes).

Detailed figures and the distribution of data confirm that the functionally reduced frequencies of each word form in my dataset did not modify the distribution of variables within the selected sample of paradigms as attested in the reference corpus. The advantage of reducing token frequencies into a range of 1-1001 is represented by a heavily scaling down of computational resources. In addition, reducing frequencies of the selected subset of verb forms for the three languages – Spanish, German and Italian – into the same range, makes them fully comparable, regardless of the different size of the reference corpora.

Out of the 50 selected Spanish sub-paradigms, 27 are irregular and 23 are regular[26].

Figure 7 shows token frequencies for the two formally defined categories, namely irregulars and regulars.

---

[26] For details, see § 4.2.
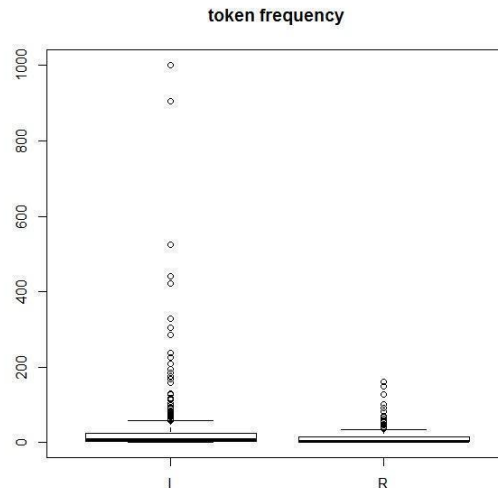
Figure 7: Token frequency distribution for irregular (I) and regular (R) paradigms in the Spanish dataset.

A one-way ANOVA test[27] shows a significant effect of (ir)regularity in frequencies distribution (p-value <0.001).

Concerning words, the average length is 7.235 (with a standard deviation of 2.33). The minimum length is two, and the maximum length is 15.

4.1.2. THE GERMAN DATASET

Out of the 50 most frequent German verb paradigms 34 are formally classified as irregulars and 16 as regulars. The average length of fully inflected verb forms is 6.356 (with a standard deviation of 1.57). The minimum length is three and the maximum length is 11.

Once again, corpus-based and functionally-reduced frequency distributions are highly correlated (r=0.99), and frequencies are not normally distributed (p-value < 0.001). Detailed figures are given in Table 3.

---

[27] Analysis of variance, or ANOVA, indicates that there are differences between two groups of data – words of irregular paradigms and words in regular ones. The strength of the assertion is quantified by giving the significance strength, i.e. p-value. For each p-value that is lower than 0.05 the hypothesis is accepted.

GERMAN SET	Corpus frequencies	Dataset frequencies
Mean (standard deviation)	564.3 (2159.77)	13.71 (48.69)
Minimum	0	1
Maximum	44361	1001
1st quantile	15	1
3rd quantile	430	11
Median	160	5

Table 3: Detailed figures for the selected dataset with reference corpus frequencies (left column) and the functionally adjusted frequencies (right column)

A one-way ANOVA test shows a significant effect of (ir)regularity in frequencies distribution (p-value <0.001). See Figure 8 for frequency distributions in the two formal categories.
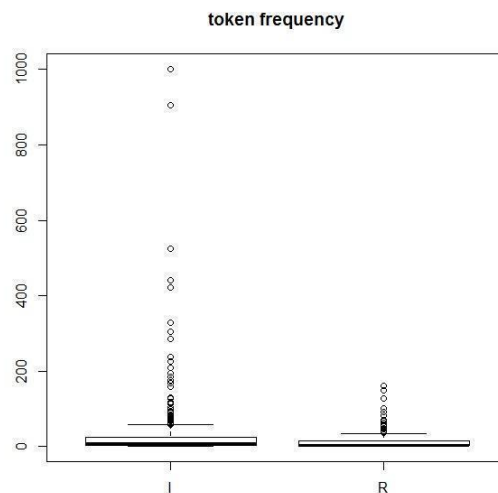
Figure 8: Token frequency distribution for irregular (I) and regular (R) paradigms in the German dataset.

4.1.3. THE ITALIAN DATASET

With the same criterion the Italian set of the 50 most frequent verb paradigms has been selected. As for Spanish and German, corpus-based and functionally-reduced frequency distributions are highly correlated (r=0.99), with the frequencies not normally distributed (p-value < 0.001). See Table 4 for detailed figures.

ITALIAN SET	Corpus frequencies	Dataset frequencies
Mean (standard deviation)	14975 (103994.9)	6.936 (41.38)
Minimum	0	1
Maximum	2514092	1001
1st quantile	24	1
3rd quantile	8243	4
Median	1166	1

Table 4: Detailed figures for the selected dataset with reference corpus frequencies (left column) and the functionally adjusted frequencies (right column).

The average length of Italian verb forms is 7.082 (with a standard deviation of 1.88). The minimum[28] length is two and the maximum length is 11. 27 paradigms are formally classified as irregulars and 23 as regulars. Once again, the token frequency is differently distributed depending on (ir)regularity. In fact, a one-way ANOVA test shows a significant effect of (ir)regularity in frequencies distribution (p-value <0.01). Figure 9 shows frequency distributions in the two formal categories.
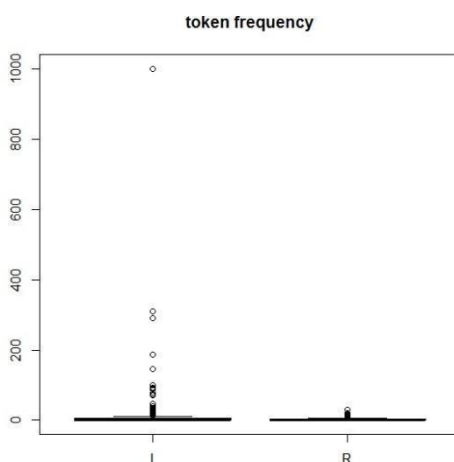
Figure 9: Token frequency distribution for irregular (I) and regular (R) paradigms in the Italian dataset.

---

[28] The third person singular for the simple present of *essere* (be), *è* (is), is encoded as *e'*. Thus the input to the TSOM is *#,E,',$*.

4.2. Spanish dataset annotation

For each fully inflected verb form in the Spanish dataset, stem, suffix, and prefix if any, are annotated by defining their length in terms of the number of symbols. The segmentation has been carried out by separating the endings – representative of the three Spanish conjugations, *-ar, -er, -ir* – from the stems. Spanish verbal inflection does not present any prefix in the past tense forms. Some difficulties have been experienced due to irregularities of certain inflection pattern, as well as to occasional phonological alternation within the stem.

Spanish verbs are classified as either regulars or irregulars by *Nueva gramática de la lengua española* (RAE, 2011: 57-69). Whereby verb paradigms did not fit the canonical inflectional patterns of *amar, temer, partir* (RAE 2011: 57), a more detailed analysis was necessary.

Spanish verbal inflection is subjected to different degrees of irregularities, which can be roughly divided into vowel, consonant, or mixed change. Vowel change leads to the so called *diphthongisation*, when the stem vowel is replaced by a diphthong, as follows: *o - ue*, like in *contar - cuento*; *e - ie*, like in *pensar - pienso*. Consonant change produces similar output, resulting into an alternation of a consonant, e.g. in *hacer - hago*. Both strategies are comprised in the mixed irregular pattern. However, the highest level of irregularity is represented by suppletive roots, which corresponds to a root alternation. This is the case of *ir – voy – fui* ("(to)go" – "(I)go" – "(I)went"), *ser – soy – fui* ("(to)be" – "(I)am" – "(I)was"), among others.

Another type of irregularity of Spanish verbs is dictated by the vowel alternation, known as "raising", because the stem vowel "raises" from "e" to "i". This phenomenon belongs to the third conjugation, e.g. *pedir – pido*, and sometimes it also integrates the diphthongisation, like in *mentir – miento* (Embick, 2012).

As already mentioned, each verb form was annotated with its stem and suffix (i.e. an inflectional ending), corresponding to the Data-Set columns "root length" and "suffix length". Some verb forms show peculiarities in the inflection pattern, both

within the stem and the ending. For this reason, it seems necessary to explain the reasons of the morphological segmentation adopted in this study. First of all, some orthographic variations are due to different phonological contexts (Zamorano Mansilla, 2008), which are completely predictable, like in *utilizo - utilicé*, where *z* and *c* correspond to the same phoneme /θ/.

Where endings are not directly linked to the stem, e.g. *seguir – sigue*, I adopted the same criterion as for regular forms, that is to say, solely the suffixes were isolated from the rest of the verb-form, so that the exceeding vowel was considered as part of the stem.

One more category of irregular verbs in Spanish exhibit just one irregular inflected form, usually the first singular of the Present Indicative, e.g. *poner - pongo, salir - salgo, and hacer - hago*, although in these verbs almost all the other selected forms are completely irregular.

4.3. Experimental design

A battery of experiments is designed to investigate the functional behavior of TSOMs, and their morphological organization, when trained on different lexica in different training conditions.

To simulate different conditions of input exposure to more than one language, and to address issues of L2 acquisition and processing and L1/L2 competition, an incremental training regime was adopted. That is, with no resetting of the TSOM parameters and no loss of already stored information, TSOM maps are exposed to the selected datasets in various combinations:

(i) one condition of Spanish L1 and German L2, that is the set of 750 inflected forms for Spanish language is shown to the map for 100 epochs (in the range of learning epochs 1-100), and the set of 750 German forms are

shown for 50 epochs (in the range of learning epochs 51-100[29]);

(ii) the reverse condition, i.e. German L1 and Spanish L2;

(iii) one condition of Spanish L1 and Italian L2, that is the set of 750 inflected forms for Spanish language is shown to the map for 100 epochs (in the range of learning epochs 1-100), and the set of 750 Italian forms are shown for 50 epochs (in the range of learning epochs 51-100);

(iv) the reverse condition, i.e. Italian L1 and Spanish L2;

(v) a perfect bilingual experimental condition, with Spanish and German as both L1 (i.e. both datasets are input in the range of learning epochs 1-100);

(vi) the perfect bilingual experimental condition, with Spanish and Italian as both L1 (i.e. both datasets are input in the range of learning epochs 1-100);

(vii) 3 strictly monolingual contexts – Spanish, German and Italian – for the 3 experimental conditions where only one dataset is shown, to compare results with bilingual regimes.

For each training condition, word forms of the 3 datasets are input according to their token frequencies, and with an identical parametrical set 5 TSOM instances are repeatedly run so to average results[30]. Each is represented by a time-series of symbols, namely a sequence of orthographic letters, administered one symbol at a time.

The choice of the orthographic code is determined by a better availability of corpora. In addition, Spanish, German and Italian can all be defined as substantially transparent languages, that is with a transparent orthography[31].

---

[29] An exposure to L2 starting from epoch 51 (in a learning range 1-100) cannot be defined as an early bilingualism. However, the richness of L2 input – both in type and token frequency (i.e. richness of the vocabulary and amount of exposure) – brings to high accuracy results at the end of learning. These may be thought as corresponding to a good proficiency in mastering L2 lexicon, although some differences in the overall organization must be noticed.

[30] The bigger the number of instances, the more reliable the results, since averaged results minimize randomness factors. When defining the amount of repetitions of the same configuration, the overall computational time must be taken into account. The experiment at issue is quite time-consuming, in fact for each combination, for each dataset, around three hours were devoted to complete the training. The estimation, in this case, for all language combinations, results in about 225 training hours.

[31] Spanish and Italian are fully transparent; German is transparent to a lesser extent than Spanish and Italian.

A whole word is presented to a TSOM starting with a start-of-word symbol (#) and ending with the symbol ($) representing the end of a word. Each symbol is coded as orthogonal to any other symbol (see symbol encoding vectors, Annex 1).

For the sake of precision, it must be pointed out that the Spanish and German monolingual maps are 40 x 40 memory node maps, whereas for any other training regime lexica are shown to 42 x 42 node maps. The estimation is based on the overall complexity of lexical input, that is in consideration of the average length of input words, formal redundancy (i.e. recurrent morphological structure shared by word forms, and number of word types). It must be noticed that, for example, in the three sets of 750 words, there are 715 word-types for Spanish, 504 types for German, and 748 for Italian. For the selected 15 cells, there are, in fact, many homographs in German, some homographs in Spanish, and only four in Italian[32].

After having outlined the method and the datasets used for the study, I will report in detail on some experimental results in the following Chapter.

---

[32] In the Italian dataset we have *sono* for *io sono* 'I am' and *essi sono* 'they are'; *stato* as the past participle for *essere* 'to be' and *stare* 'to stay'. German verbal inflection is strongly characterized by redundancy: infinitive, 1st and 3rd plural persons for the indicative present share the same form, both in regular than in irregular predictable paradigms, as well as 1st and 3rd singular persons for the simple past (präteritum), and 1st and 3rd plural person for the simple past. Idiosyncratic paradigms do not follow this redundancy.

5. EXPERIMENTAL RESULTS

After training, the memory content of each of the trained TSOM maps were tested to verify the internal organization on two tasks: word RECODING and word RECALL. The task of RECODING consists in quantify the accuracy of the map's activation on input forms. An input word is recognized correctly if each best matching unit (BMU) in the activation chain is correctly associated with the current input symbol. Errors are counted when an input symbol activates a BMU associated with a different symbol. In this case, the whole word is considered wrongly recoded. It can be said that word recognition depends on the current input stimulus, that is, as a measure of short-term memory.

Conversely, word RECALL depends on the long-term memory of a map, and simulates the process of retrieving a sequence of symbols from its BMUs. Errors occur when the map misrecalls one or more symbols in the input string, by replacing it with a different symbol or by outputting correct symbols in the wrong order. Also for this task, errors on one symbol only are counted as an error (Marzi et al., 2014a).

Results on recoding and recall tasks for each language in the different learning conditions are averaged over the five repetitions and are given in Table 5.

Language – learning epoch	Recoding accuracy	Recall accuracy	standard deviation
Spanish monolingual - 100	100%	99.30%	0.61
German monolingual - 100	100%	99.52%	0.27
Italian monolingual - 100	100%	98.69%	0.55
Spanish L1 - 50	100%	98.88%	0.56
Spanish L1 - 100	100%	98.38%	1.23
German L2 - 100	100%	91.98%	1.92
German L1 – 50	100%	99.33%	0.36
German L1 – 100	100%	98.65%	0.51
Spanish L2 – 100	98.43% (0.88 sd)	92.67%	0.87
Spanish L1 – 50	100%	99.33%	0.44
Spanish L1 – 100	100%	99.22%	0.47
Italian L2 – 100	100%	95.24%	1.26
Italian L1 – 50	100%	99.06%	0.34
Italian L1 – 100	100%	96.39%	0.61
Spanish L2 - 100	100%	97.03%	1.07
Spanish L1 – 50	100%	97.31%	0.55
Spanish L1- 100	100%	97.43%	0.47
German L1 – 50	100%	97.98%	0.66
German L1 - 100	100%	98.17%	0.55
Spanish L1 – 50	100%	98.07%	0.46
Spanish L1 – 100	100%	96.07%	1.69
Italian L1 – 50	100%	98.07%	0.46
Italian L1 - 100	100%	96.15%	1.65

Table 5: Recoding and recall accuracy are given in percentage for each language in all training conditions. Scores are averaged on 5 repetitions.

5.1. *Monolingual regime*

A detailed analysis started from the acquisition time of word forms for the three lexica in the three monolingual training conditions.

In Figure 10, the time course of word acquisition is given for Spanish, German and Italian monolingual training. In detail, recoding and recall are monitored over time[33].

---

[33] In detail, acquisition time is monitored, and plotted accordingly, over the following learning epochs: 1:30, 50, 51:61, and the final epoch, 100.
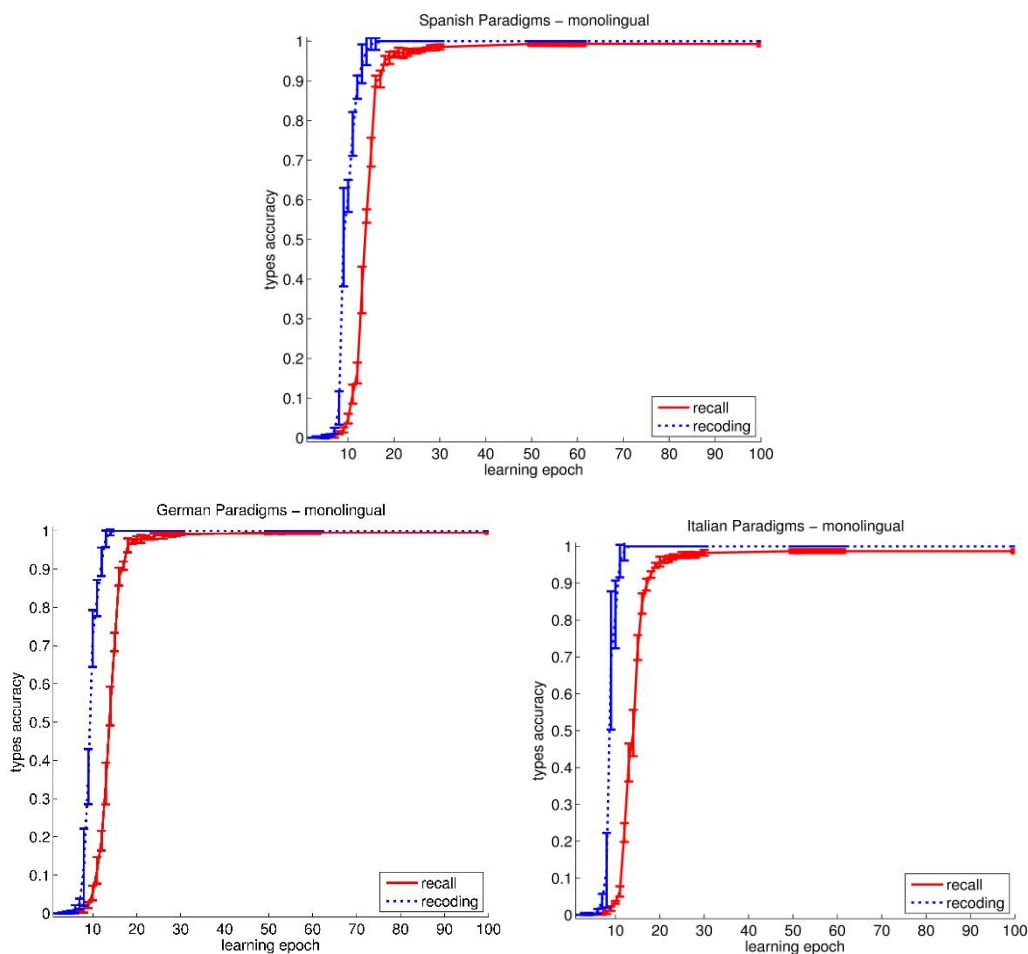
Figure 10: Recoding and recall accuracy over learning epochs for Spanish (top plot), German (bottom left plot) and Italian (bottom right plot) words. Scores are averaged on 5 repetitions.

Figure 11 shows the acquisition epoch for the 50 most frequent paradigms in three strictly monolingual contexts, going from the fastest – namely the easier to learn - paradigm on the top (*ver* - *restare* - *werden*) to the most difficult paradigm on the bottom. The paradigm acquisition epoch provides an estimate of the average time necessary for all forms of the paradigm to be recoded and recalled correctly.

It shall be observed that regular and irregular verbs are mixed, which reflects the balance between the facility to learn regular patterns and the importance of frequency in the acquisition of irregular patterns. In addition, it should be noticed that in case of one or more forms representing a challenge for storage and

retrieval from memory storage, the whole paradigm acquisition time will result delayed. It is the case, for example, of *arbeiten* in German, and *stare* in Italian, where regular past forms represent a challenge for correctly retrieving their memory traces due to the repetition of chunks (*arbeit-et-et*, *ste-ste*).
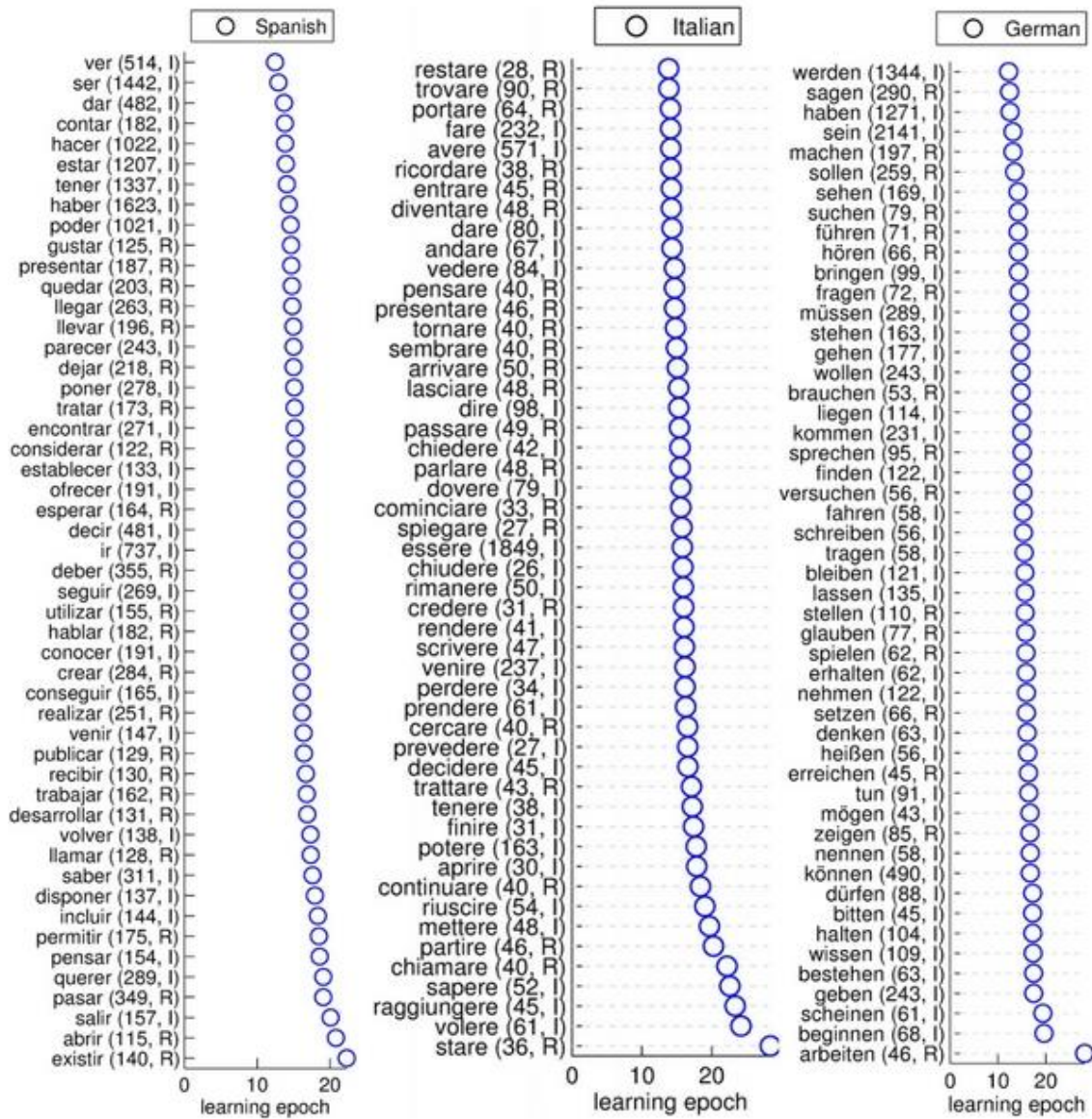
Figure 11: Spanish, Italian, and German data set course of acquisition, where paradigms are ranked by increasing learning epoch (from top to bottom).

The simulation carried out by TSOMs can be thought of as if a child was gradually acquiring a lexicon, starting from the most frequent and short units, then learning increasingly longer and less frequent ones (Marzi & Pirrelli, 2015). For instance, surface word relations lead to the morphological organization in the lexicon. A gradient of perceived regularity-irregularity, rather than regularity vs. dichotomy, plays an important role in the formulation of the morphological organization to the extent to which regularity/predictability facilitates the emergence of common patterns, whereas highly irregularity/unpredictability makes it more difficult.

In fact, regular–irregular distinction is not as categorical as it is commonly held. Degrees of irregularity affect language acquisition differently. This observation should become evident by looking at the Spanish paradigm time of acquisition (Figure 11, left panel), where the first nine verbs are all irregulars. It should be appreciated that these nine verbs exhibit different degrees of irregularity. In fact, the less irregular ones have lower cumulative frequency (reported in brackets)[34], meanwhile the highly irregular ones are attested with very high frequency. This observation confirms the assumption that the frequency factor can neutralize the complexity of some irregular patterns, by supporting the acquisition of word forms in isolation. In this case, the situation is diametrically opposed for regulars: among the last forms appear some regular verbs, such as *permitir*, *pasar*, and, most importantly, *abrir* and *existir*, the latter two being in the very last two positions. Despite their regularity, these forms are on the bottom of the list because of their scant frequency. The case of *abrir* is especially interesting, because although the paradigm is classified as regular, this verb has an irregular Past Participle *abierto* that certainly does not support a faster learning. In this, and similar cases, in fact, the whole paradigm can benefit from the cumulated stem frequency of all forms to a lesser extent. This is the overall dynamic that supports some less attested paradigms (see for example *gustar*, *presentar*, *quedar*, *llevar*).

---

[34] The cumulative frequency refers to the sum of the frequencies for all the selected 15 forms of the verb.

For German and Italian, the paradigm acquisition times confirm that there is not a unique factor determining the acquisitional dynamics. Neither word frequency nor regularity vs irregularity by itself can explain the evidence. All these factors must be considered in their complexity. Moreover, word frequency is not the alone quantitative amount to be taken into account, since the distribution of token frequencies among forms in the same paradigm plays a significant role.

In this perspective, previous studies (e.g. Marzi & Pirrelli, 2015: 519) posit that skewed frequency distributions tend to slow down the acquisition of the paradigm as a whole. Since the paradigm acquisition epoch is calculated as the average acquisition time of all forms, less-frequency items will affect it negatively. This important issue can be illustrated with the form *arbeiten*, showing a significant acquisitional delay with respect to the other 49 German verbs, which can be attributed to both its skewed frequency distribution and the length of some of its forms[35].

Particular cases apart, the paradigm frequency has less effect on regular verbs, whose acquisition is facilitated by the neighboring family of words, rather than by frequency as in the case of irregulars. The regular forms, in fact, can not only be inferred from other forms of the same paradigm, but they can benefit from a sort of boosting effect given by the cumulative frequency of inflected forms that share the same stem (Marzi et al., 2014a).

As already mentioned in §3.3.3, words get coded in our long-term memory as activation patterns of time-series of symbols. When receiving an input stimulus at time *t*, TSOM triggers the competitive activation of all nodes. This dynamic of lexical acquisition and processing is displayed in Figure 12, where concurrently stored words result in a partially overlapping activation pattern between *abrir - abierto*. The memory grid[36] shows here TSOM's both internal synaptic

---

[35] Within the paradigm of *arbeiten*, *arbeitest* and *arbeitetest* are hapax forms and, interestingly enough, they also belong to the group of the longest units. Other forms, such as *arbeitetest* and *arbeiteten*, share the same characteristic of sequence repetition, in addition to their low frequency.
[36] The size of memory grid is here 40 X 40 nodes.

connectivity between sequentially activated nodes (see arrows in Figure 12) and levels of co-activation (see color for each node).
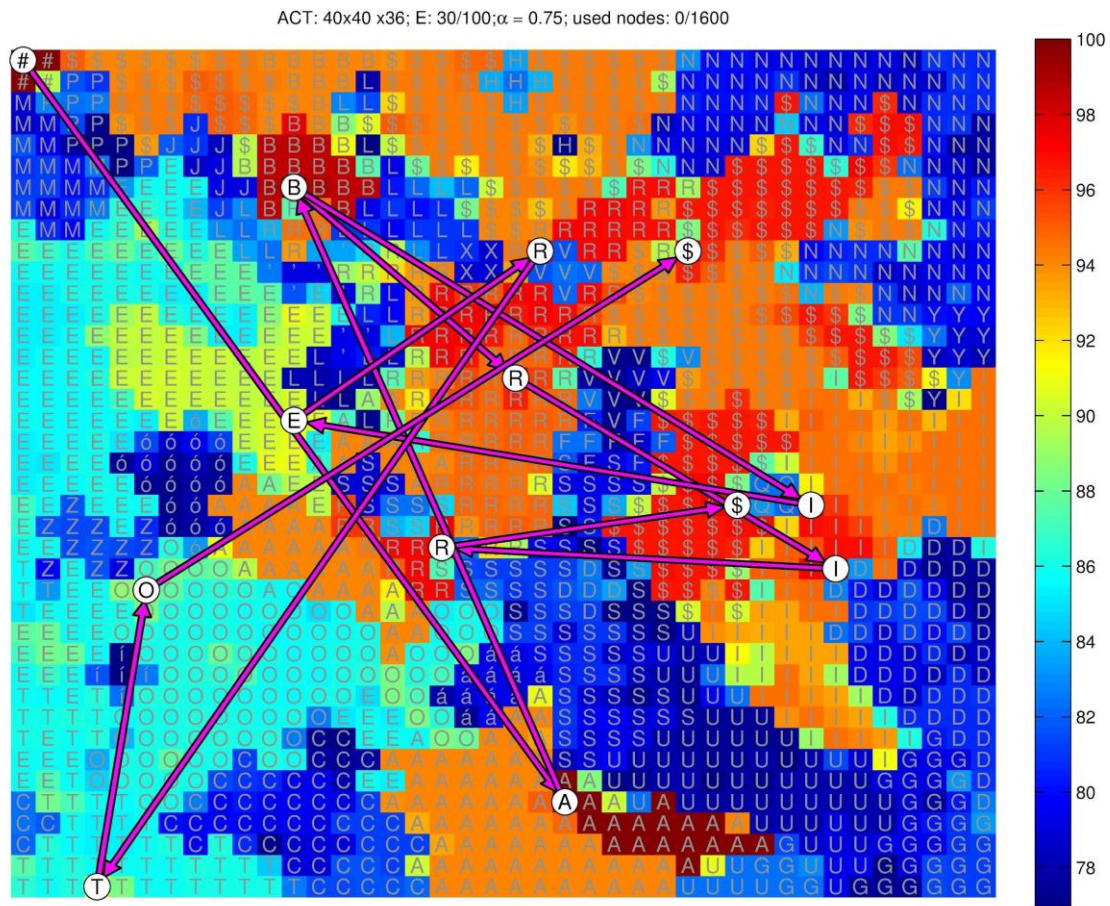
Figure 12: Co-activation levels for chain of memory nodes that activate *abrir* - *abierto*.

The mapping of colors allows for the reduction of the representation of multidimensional data on two-dimensional grid of nodes (where red stands for maximum activation, and blue for minimum activation). If we follow the activation pattern from the symbol # (start of word) to the symbol $ (end of word)[37], we'll notice that the pattern branches off in correspondence of the node *b*, which is the last shared node between the two forms.

---

[37] *Abrir* and *abierto* are shown to the map as the input strings *#a,b,r,i,r$* and *#a,b,i,e,r,t,o$*, respectively.
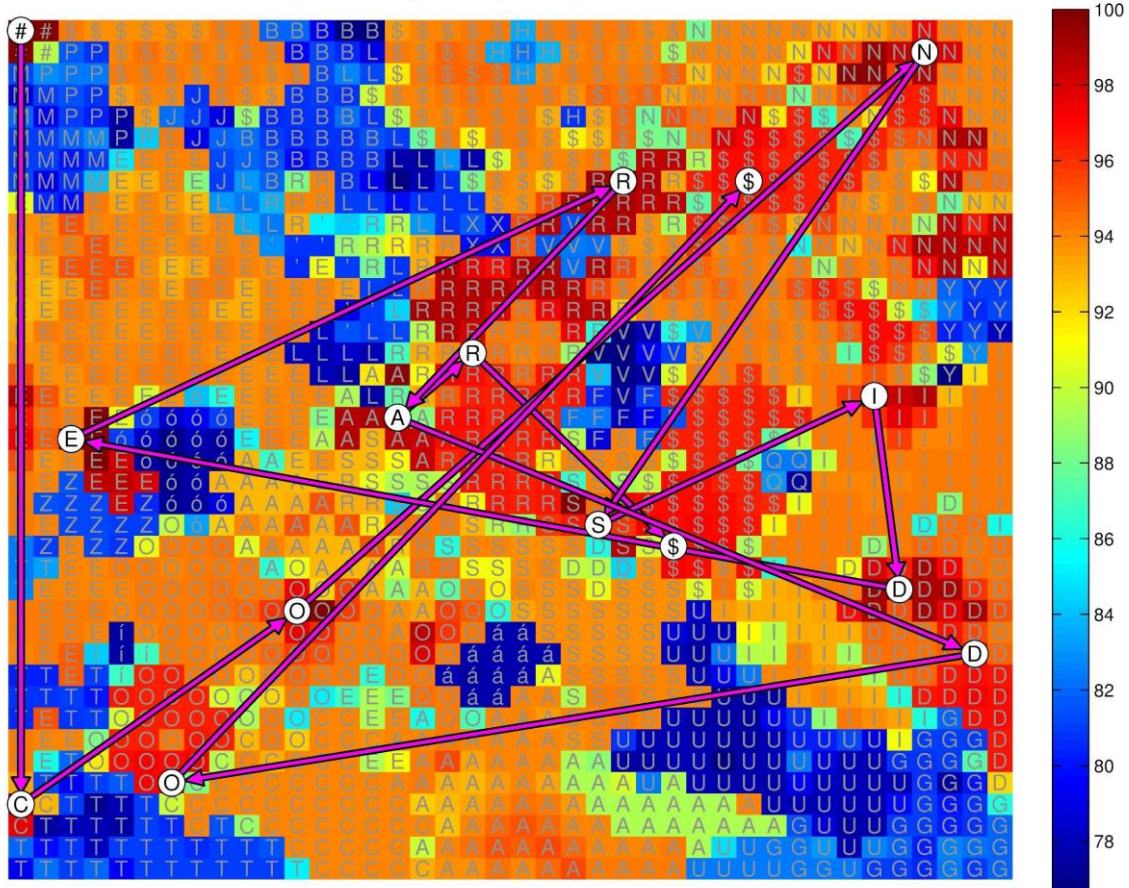
Figure 13: Co-activation levels for chain of memory nodes that activate *considerar - considerado*.

In this way, the correlation between the incoming stimuli and the already stored forms provides evidence to our assumption that storage and processing are two sides of the same coin. As further proof, Figure 13 shows the co-activation of *considerar - considerado* patterns. Once again, the overlapping nodes *c,o,n,s,i,d,e,r,a* exhibit high level of co-activation (intense red clusters of nodes) up to the thematic vowel, where the memory chain splits in two different memory traces.

It should be appreciated that although *abrir* and *considerar* are both regular paradigms, the irregular past participle form of *abrir* makes this form less co-activated to the infinitive form, for example, than *considerar* and *considerado*.

Although it has been affirmed that regularity/irregularity is a gradient of predictability, it is in any case interesting to monitor the different acquisition pace for regulars and irregulars (as formally classified).
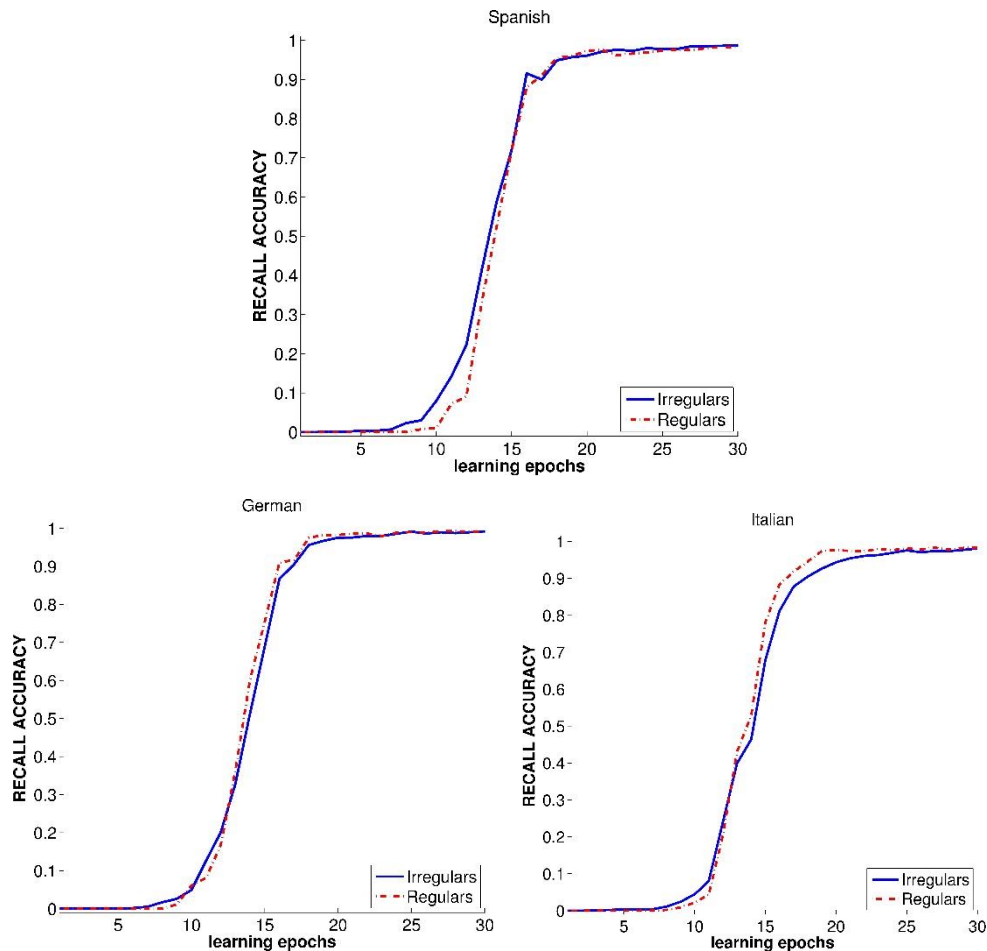
Figure 14: Learning differences between irregulars and regulars for Spanish (top panel), German (bottom left panel) and Italian (bottom right panel). Accuracy is given for word types.

Figure 14 shows for the three monolingual contexts the difference in acquisition pace between regulars (dashed line) and irregulars (solid line). Recall accuracy refers to the capacity of the network to correctly retrieve the stored form. Apparently, the difference between regulars and irregulars in the three languages is very small. Furthermore, as observed in the present experiments, the first 30 learning epochs play a crucial role, because they represent the time span when

the learning process achieves and stabilizes the highest recoding and recall accuracy.

This dynamic can be explained by verifying how the plasticity of the TSOM map lets adapt the spatial clustering firstly to the most frequent input symbols, specializing different nodes in the same cluster for symbol occurrences in different temporal contexts, and then progressively to the less frequently ones. This adaptive self-organization reaches a stable point at around learning epoch 20. In fact, there are no differences in clustering between learning epoch 20 and 30 (see Figure 15 below).
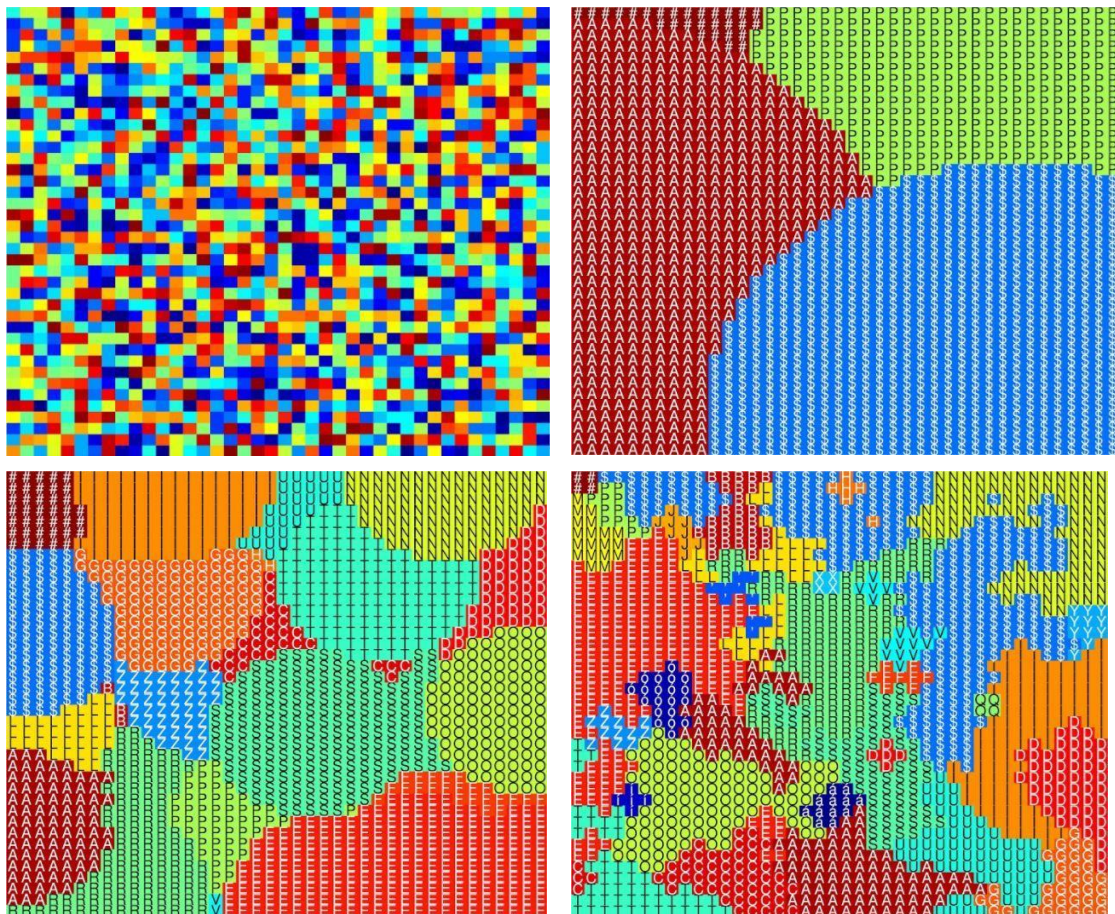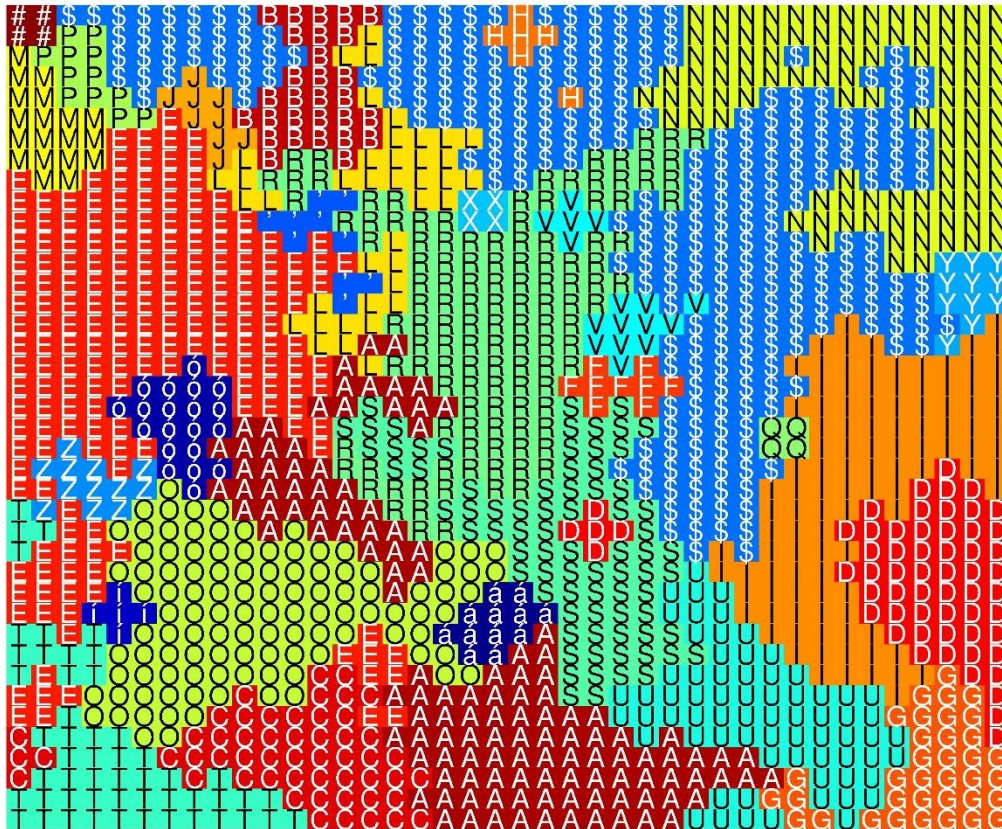
Figure 15: Symbol clustering in the Spanish monolingual regime: learning epochs 0 (top left panel), 5 (top right panel), 10 (middle left panel), 20 (middle right panel), 30 (bottom panel).

Turning back to the acquisition pace for the three languages (Figure 14), Italian and German show more similarity between them, compared to Spanish results: in both contexts irregulars are more rapidly learnt until the epoch 12 approximately, where the situation overturns. In Spanish the irregular forms lead the learning process (note the gap around epoch 10). This is due to the distribution of frequencies. In fact, it must not be forgotten that irregular forms are more frequent than regular ones (as highlighted in § 4). This is confirmed when instead of monitoring recall accuracy over word types, we observe recall accuracy over tokens (see Figure 16). In such case, differences get more evident, and confirm that irregulars are acquired earlier since they may rely on higher token frequencies of word forms.
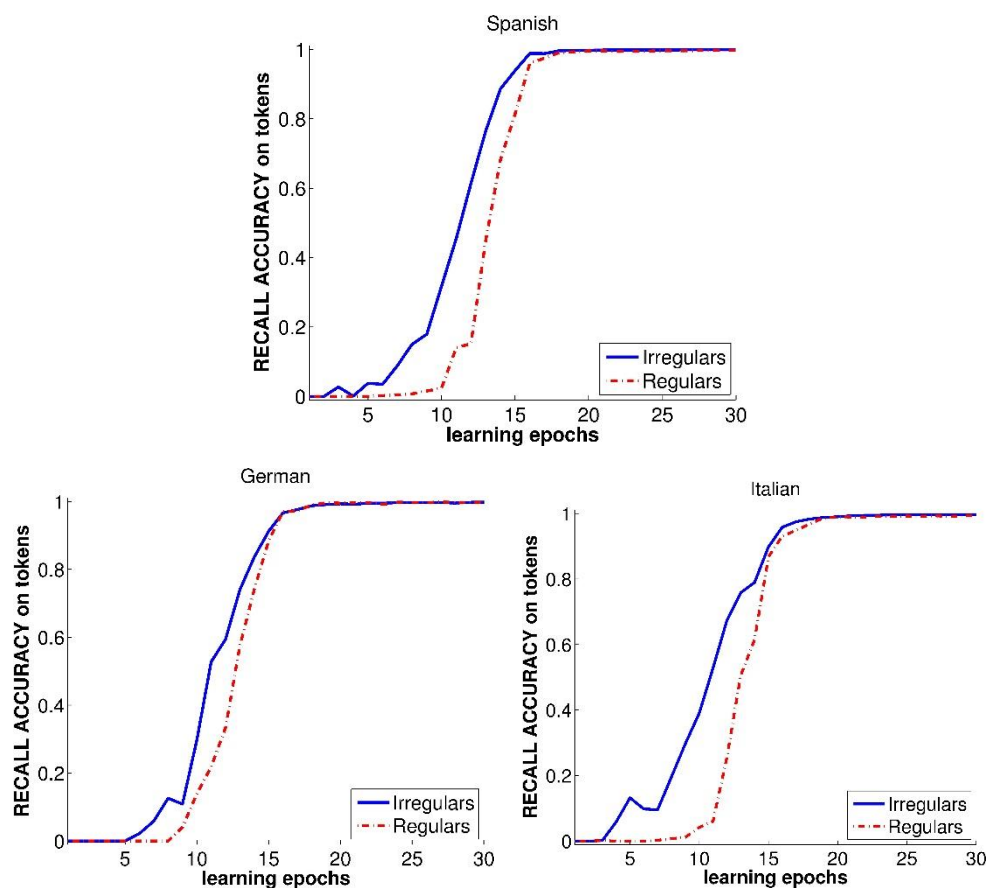
Figure 16: Learning differences between irregulars and regulars for Spanish (top panel), German (bottom left panel) and Italian (bottom right panel). Accuracy is given for word tokens.

Evidence as shown in Figure 16 supports the position that token frequency should be taken into account, since the map (in the present approach) is exposed to as many inputs as the overall amount of tokens. Therefore, it is not surprising that, on average, forms of irregular paradigms strongly benefit from their higher frequency support, compared to regulars.

The importance of type/token frequency has been emphasized by Bybee within the framework of the Network Model (see §3.2.3). In fact, the lack of processability of some idiosyncratic verbs is balanced by their very high token frequency. Figure 16 reveals this difference in frequency effect, by highlighting the huge gap between regular vs. irregular paradigms recall. The bigger difference of this kind is observed in Italian, where irregulars show a remarkable peak in correspondence to the fifth epoch. Conversely, due to their lower token

frequency, regular forms need more training to achieve a better recall with respect to the other two languages.

Preliminarily, I conclude that morphology acquisition is determined by word token frequency as well as by formal redundancy, that is to say, by morphological regularity intended as shared patterns. The more verb inflected forms share part of their superficial forms (i.e. the stem), the better and easier the acquisition of them. It is important at this stage of analysis to move to bilingual learning regimes, where it will be interesting to monitor a somewhat different behavior for L2 forms.

5.2. *Bilingual regimes*

At the beginning of this chapter, in table 5, I summarized accuracy percentage for each combination of lexical exposure. However, it is more informative to monitor the pace of acquisition during learning epochs, as already pointed out for the monolingual condition.

In detail, I consider the Spanish-L1 German-L2 regime, where the set of 750 inflected Spanish forms is shown to the map for the total amount of 100 epochs, and the set of 750 German forms are shown for a total of 50 epochs, starting from learning epoch 51 up to 100. In comparing Spanish monolingual and bilingual context, the trend is fully comparable until the epoch 51, when the map is for the first time exposed to the second language input. Figure 17 shows recoding and recall accuracy for the two lexica.
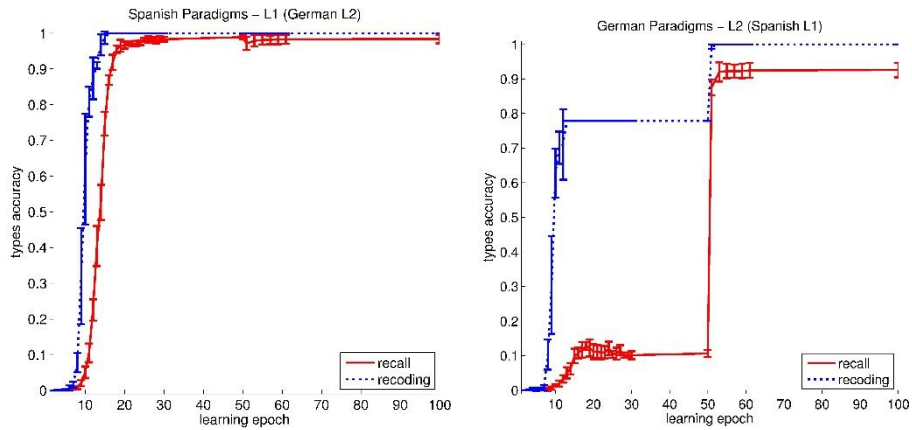
Figure 17: Acquisition pace shown as accuracy for the recoding and recall task over learning epochs for Spanish (left panel) and German (left panel) in the training condition Spanish L1 and German L2.

This behavior is due to a competition-based processing. When to the stable organization of a TSOM new lexical items are shown, in particular items of language with own specific orthotactic constraints (L2), it can be observed an initial stage where the overlapping of representations for L1 and L2 may cause an influence on L1 itself (see learning epochs 51-61). However, they are L2 lexical representations that are mostly influenced by competition, since they are also characterized by lack of context-specific specialization of its orthographic representations. A lack of fine specialization may cause processing problems in recognition and access (as observed by Bradlow & Pisoni, 1999, for lack of phonetic specialization). In fact, if an already stored word has an underspecified mental representation, the interference of strongly specified representations (typical of L1) may lead to misidentification (Cook, 2013). I contend here that this is induced by a much reduced amount of memory resources devoted to L2 specific representations.

To verify this position, Figure 18 offers a sequential representation of some stages for the bilingual regime, namely learning epochs 5, 10, 20, 50, which define the monolingual stage of acquisition for Spanish only; and learning epochs 61 and 100 defining the L1-L2 stage. Similar to what happens in the monolingual context (Figure 15), the spatial clustering prioritizes the most frequent inputs first,

finally adapting to the less frequently ones as well. Again, no difference in clustering after the learning epoch 20 is observed. In fact, there are no differences in clustering between learning epoch 20 and 50. When word forms of L2 are input to the map, some marginal nodes get recycled from L1 representations to adapt themselves to the new input. This is necessary, for example, for symbols that do not occur in L1 (in the present case umlauted vowels and sharp-s, *ä, ü, ö, ß*). See sparse nodes in Figure 18 (bottom left panel) highlighted with black circles.
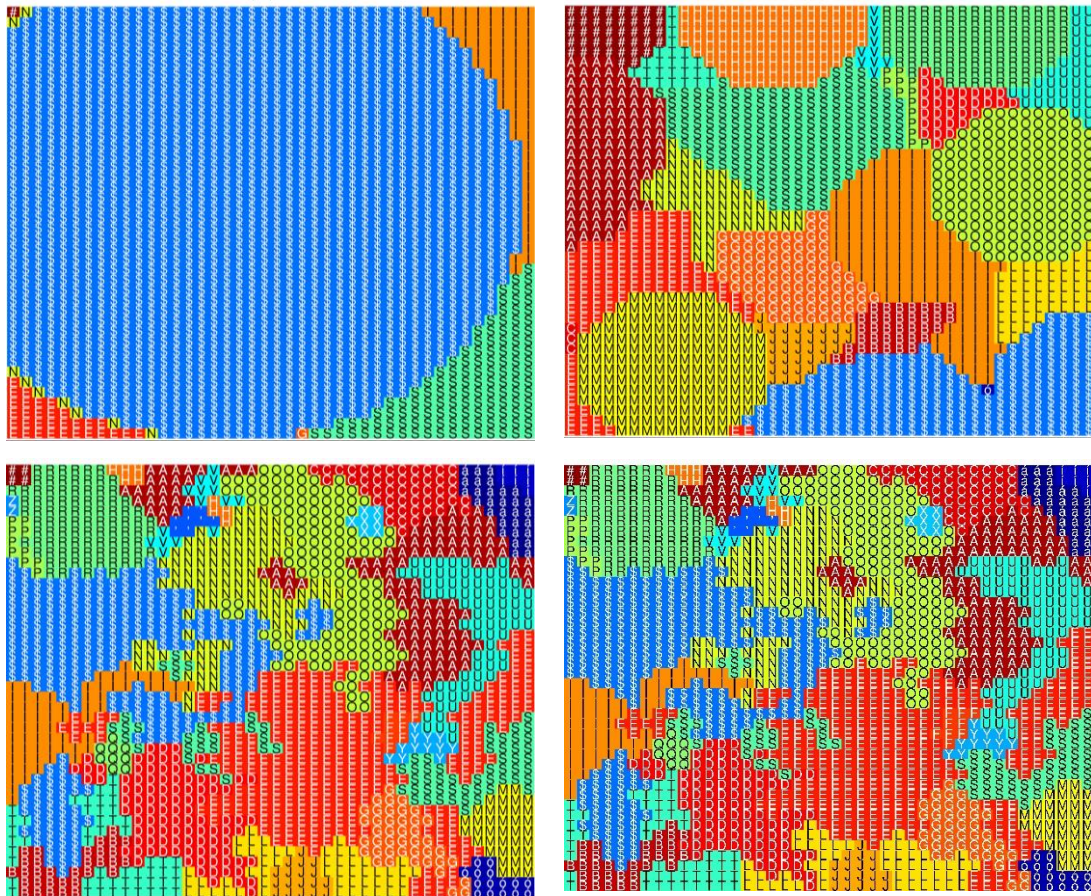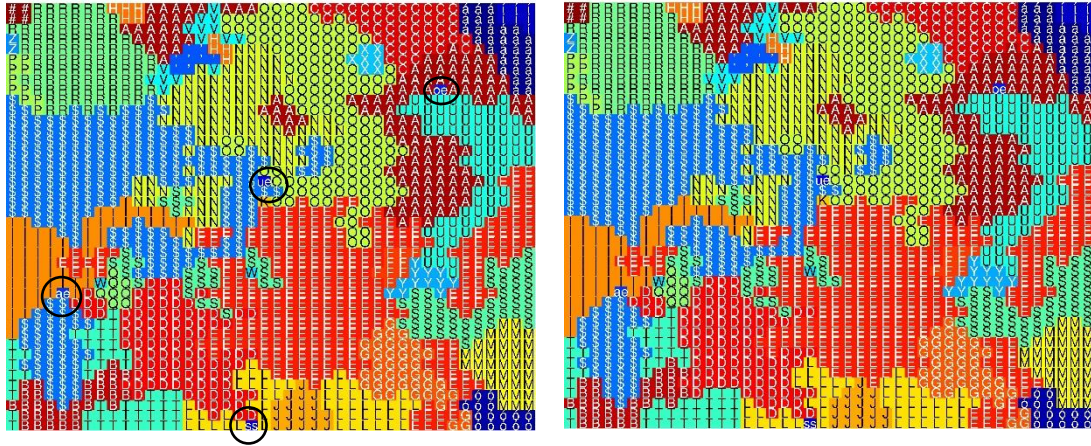
Figure 18: Symbol clustering at different learning stages in the Spanish L1 and German L2 regime: learning epochs 5 (top left panel), 10 (top right panel), 20 (middle left panel), 50 (middle right panel), 61 (bottom left panel), 100 (bottom right panel).

The recycling process of memory resources during different stages of learning can be quantitatively detected by evaluating the exact percentage of memory nodes activated by lexical representations for each language (see Table 6).

Language – learning epoch	Map nodes %
Spanish monolingual - 100	36.44%
Spanish L1 - 50	35.69%
Spanish L1 - 100	30.86%
German L2 - 100	17.07%
German L1 – 50	34.00%
German L1 – 100	28.75%
Spanish L2 – 100	17.55%
Spanish L1 – 50	36.02%
Spanish L1 – 100	33.88%
Italian L2 – 100	22.60%
Italian L1 – 50	33.56%
Italian L1 – 100	29.63%
Spanish L2 - 100	24.33%
Spanish L1 – 50	26.58%
Spanish L1- 100	26.58%
German L1 – 50	23.04%
German L1 - 100	23.04%
Spanish L1 – 50	30.67%
Spanish L1 – 100	30.69%
Italian L1 – 50	24.88%
Italian L1 - 100	24.88%

Table 6: For each training regime, percentage of specialized memory resources are given for each language.

The general trend monitored for the Spanish(L1)-German(L2) exposure condition, is also confirmed for reverse condition, as well as for the Spanish-Italian regime.

Figure 19 shows the incremental learning of the Spanish-L1 Italian-L2 regime, where the set of 750 Italian forms are added to the 750 Spanish forms starting from learning epoch 51 (up to 100). Once again, the Spanish trend for recoding and recall accuracy is comparable to the monolingual regime, with the only exception of learning epochs 51-61, that is when the map is for the first time exposed to the second language input. What is important to notice is that when the Spanish map is tested with the Italian input before learning it, that is before the L2 learning begins (epochs 1-50), a somewhat different behavior can be observed in comparison to the SpanishL1-GermanL2 regime (see Figure 17). The unseen Italian input is recoded with a higher accuracy than German input in the corresponding condition.
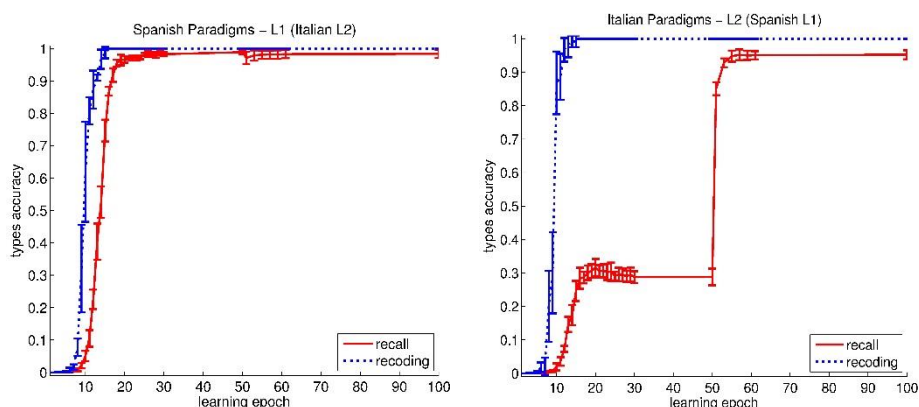
Figure 19: Acquisition pace shown as accuracy for the recoding and recall task over learning epochs for Spanish (left panel) and Italian (left panel) in the training condition Spanish L1 and Italian L2.

This is due the orthotactic likelihood that makes Spanish and Italian forms being perceived as more similar than Spanish and German ones.

70

In fact, one of the main determinant for lexical acquisition and processing is perception of similarity (see Chapter 2). Words that are perceived as similar to many other words in the lexicon may accelerate word recognition, and acquisition therefore, since wordlikeness effects interact with memory (§3.3.3).

To monitor this effect in different lexical exposures, at the end of training, it is useful to verify the TSOM's ability to anticipate a target word, that is to predict its continuation as soon as the onset is shown (Figure 20). The more symbols are anticipated, the easier the prediction of the target word.

For the sake of clarity, within the frame of child language acquisition, the best possible learning strategy should lead to generalizations extracted from the (limited) linguistic input at the disposal of the child. Consequently, from the viewpoint of the surface word, formal predictability can be considered a touchstone for lexical acquisition, inasmuch it provides a starting point by relying on the already stored information.

The point has already been made that orthotactics and redundancy play a big role in language acquisition, but what happens when two languages from different families coexist in the same bilingual context?
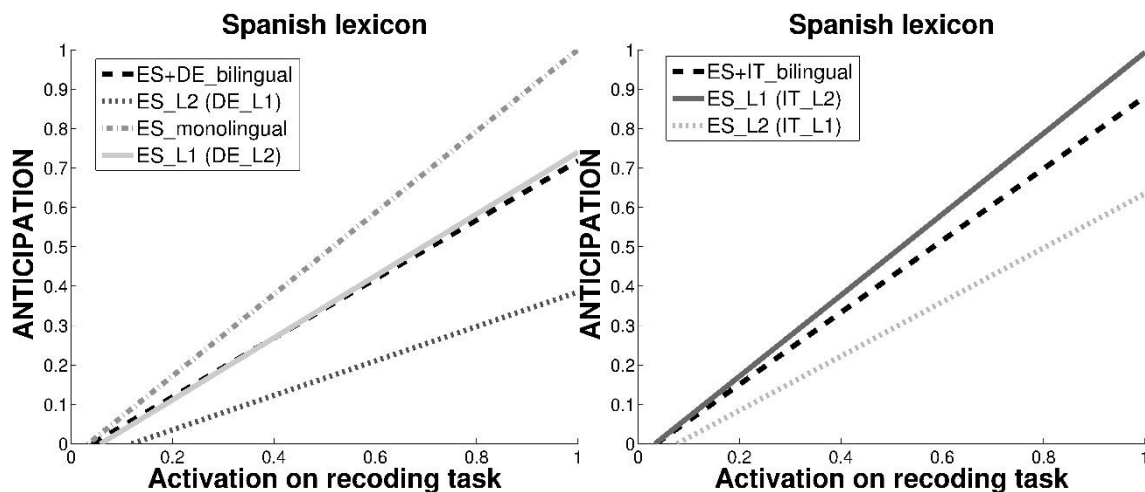
Figure 20: TSOM's ability to anticipate input words at the end of learning, by showing progressively the symbol sequences. Lines plot the anticipation over the activation levels for Spanish in the different exposure conditions: in combination with German (left plot) monolingual (grey dashed line), bilingual (dark dashed line), L1 (grey solid line), L2 (dotted line); with Italian (right plot) bilingual (dark dashed line), L1 (grey solid line), L2 (dotted line).

71

Figure 20 (left plot) shows this ability on the Spanish lexicon in the four training conditions, L1 (when German is L2), L2 (when German is L1), bilingual and monolingual ones. Figure 20 (right plot) shows the ability for Spanish in the Spanish-Italian combination. A gradient trend should be appreciated, ranging from monolingual condition to L2 training condition.

Intuitively, monolingual Spanish condition features the best anticipation ability, as no L2 is there to obstruct it. Conversely, the worst results are to observe in the situation in which Spanish is in the L2 condition. By comparing the three bilingual conditions between the Spanish vs. German and Spanish vs. Italian, the plots reveal that the latter pair offer better results in terms of word anticipation (note the slope difference between the two sets of regressions). Such difference can be attributed to the linguistic distance, that is to more similar orthographic patterns, which ultimately facilitates word processing and acquisition.

I conclude the experimental section by turning back to the paradigm acquisition epoch (as shown in Figure 11, §5.1, for the monolingual contexts) in the bilingual training conditions. It is important to observe the differences for the same dataset in the different training conditions, that is the monolingual regime (Figure 11), and the bilingual ones (Figures 21-28). As already express, the paradigm acquisition epoch provides an estimate of the average time necessary for all forms of a paradigm to be recoded and recalled correctly. In addition, the time span[38] of acquisition displays for each paradigm clear figures of the difficulty of acquisition, as a function of morphological ir/regularity and token frequency. The shorter the span, the easier the acquisition.

It should be observed that when a lexicon is in the L1 training condition, with the only exclusion of a few paradigms for which a couple of forms get never be acquired, the time span of acquisition is short, and benefit by either high token frequency of word forms, or by morphological regularity. On the contrary, when

---

[38] Paradigm acquisition span has been defined as the number of epochs it takes to complete the acquisition of a paradigm after the first member of the paradigm is acquired (Marzi et al., 2014a).

the same lexicon is in the L2 condition, the acquisition dynamic benefit from these both determinants to a lesser extent. In fact, in a lesser number of cases token frequency and formal regularity (lexical redundancy) succeed in contrasting the competition effect of the L1 lexicon.
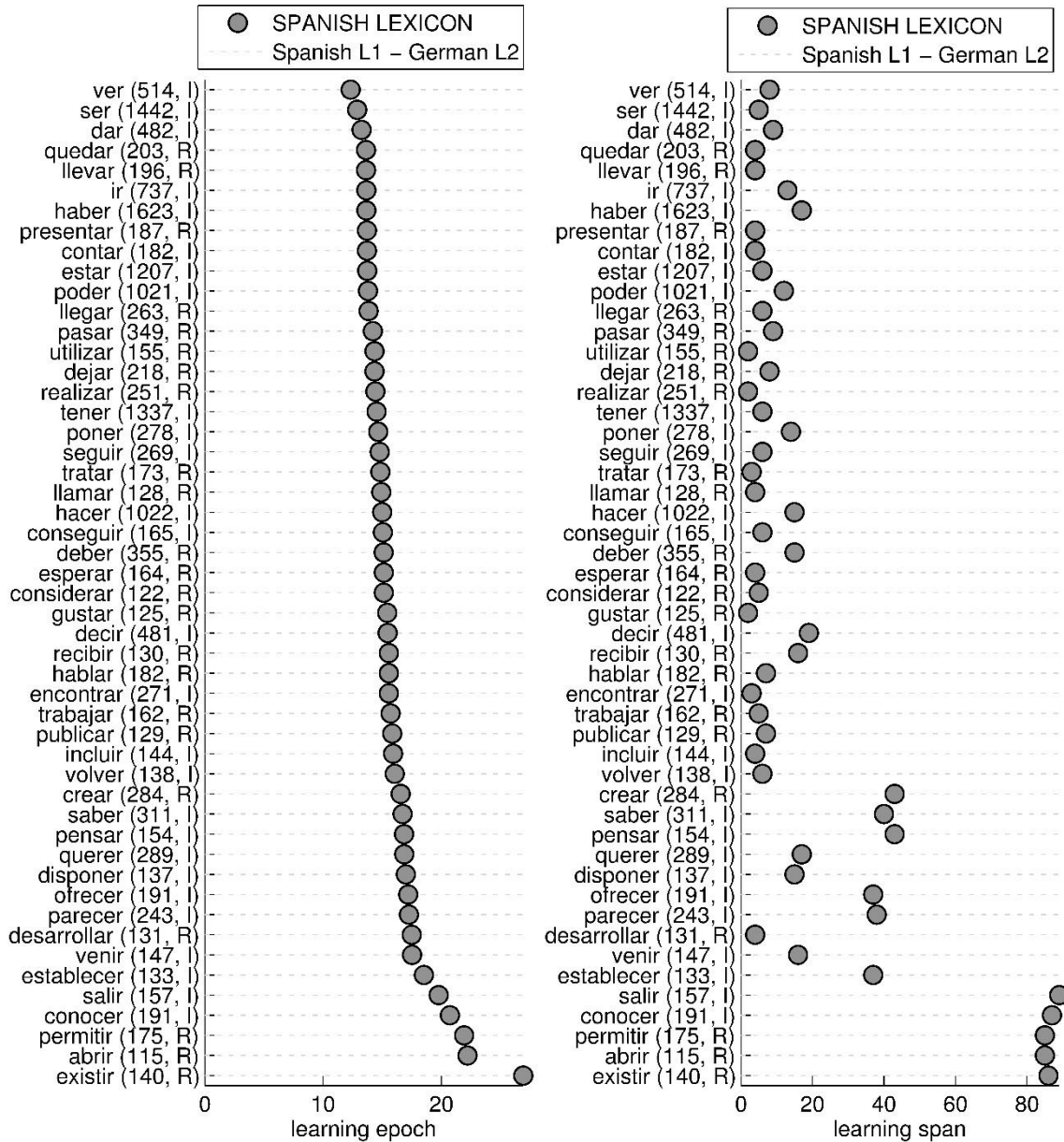
Figure 21: Spanish course (left plot) and span (right plot) of acquisition and in the Spanish_L1 and German_L2 condition, with paradigms ranked by increasing learning epoch (from top to bottom).
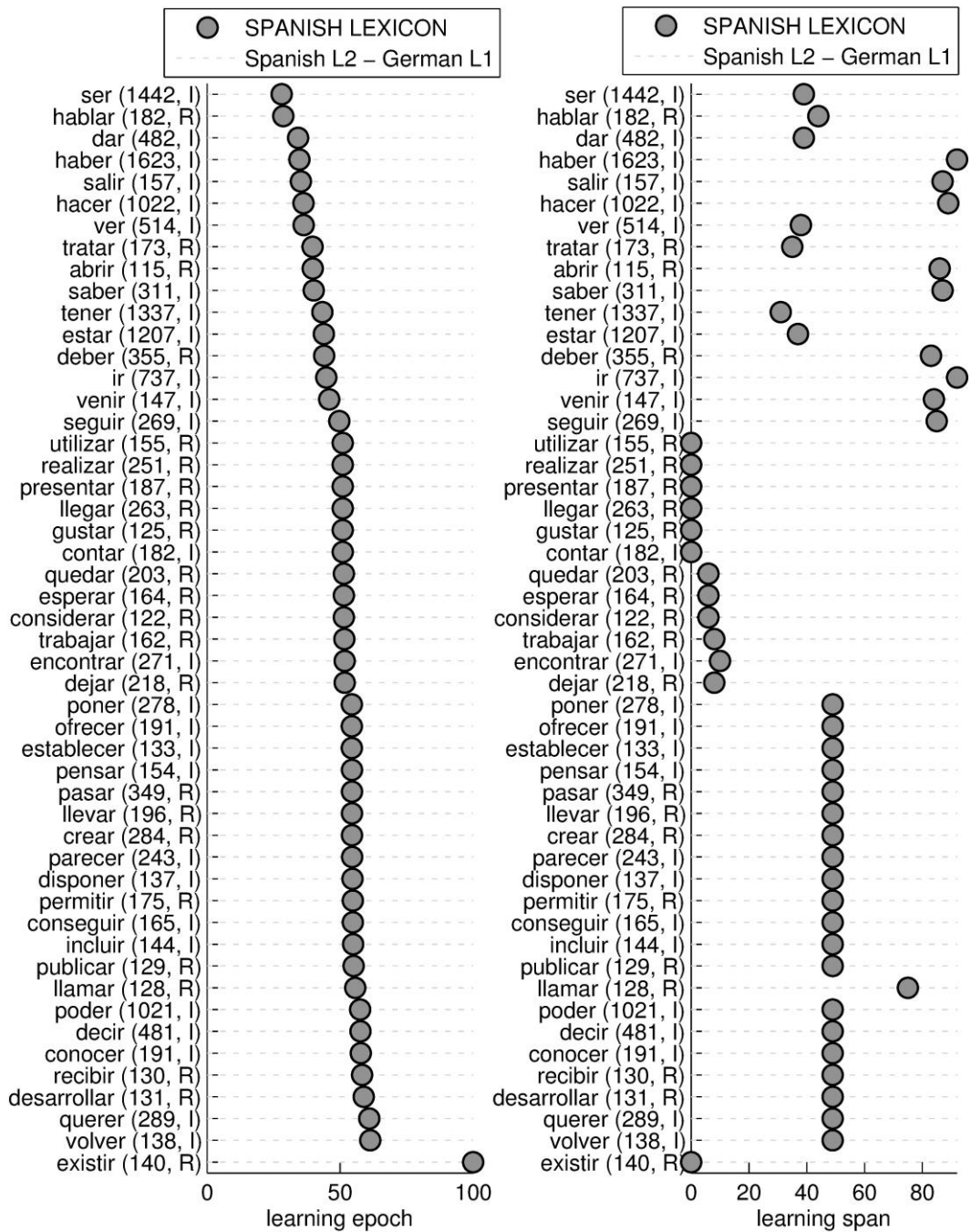
Figure 22: Spanish course (left plot) and span (right plot) of acquisition and in the Spanish_L2 and German_L1 condition, with paradigms ranked by increasing learning epoch (from top to bottom).
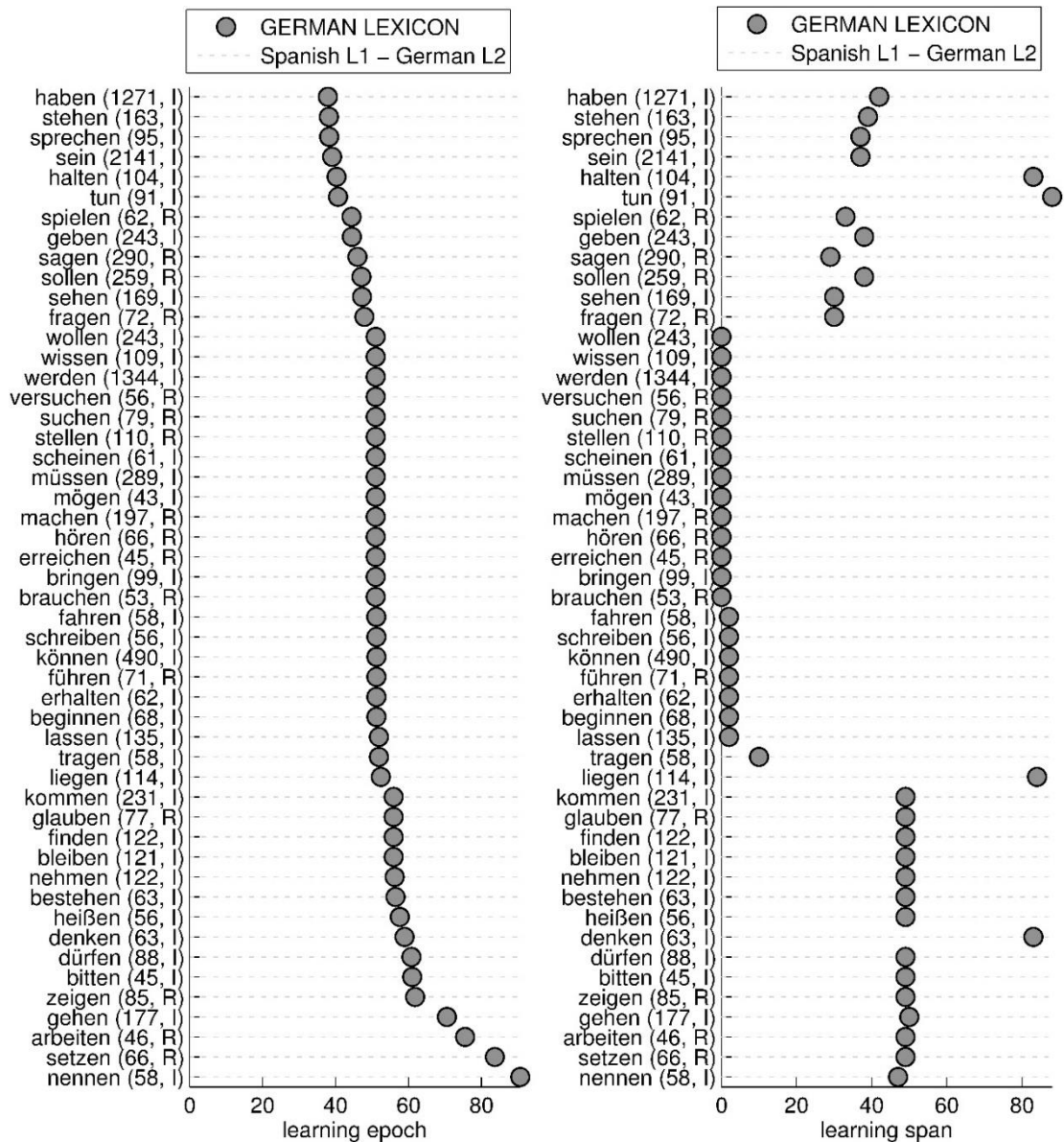
Figure 23: German course (left plot) and span (right plot) of acquisition and in the Spanish_L1 and German_L2 condition, with paradigms ranked by increasing learning epoch (from top to bottom).
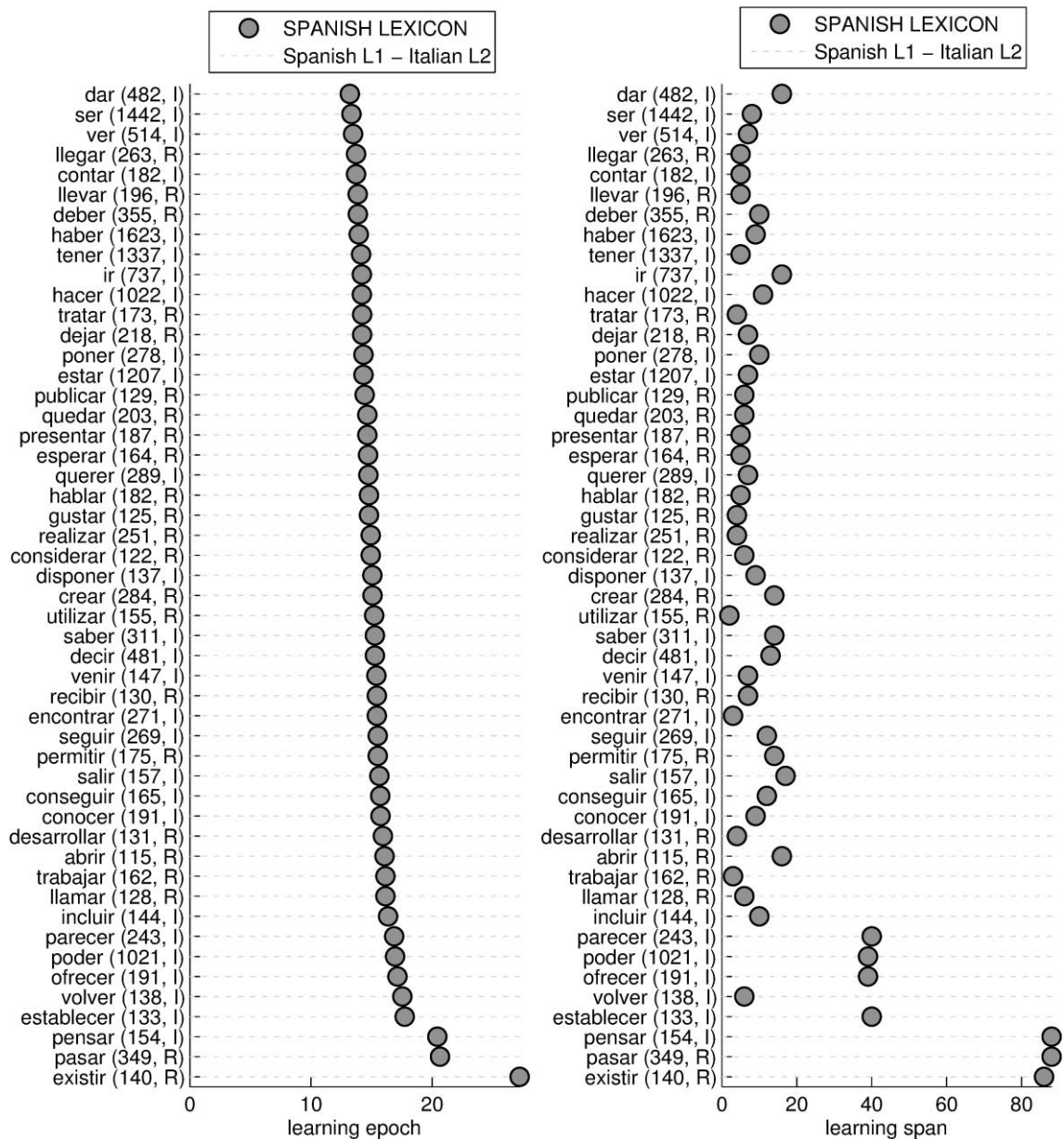
Figure 24: Spanish course (left plot) and span (right plot) of acquisition and in the Spanish_L1 and Italian_L2 condition, with paradigms ranked by increasing learning epoch (from top to bottom).
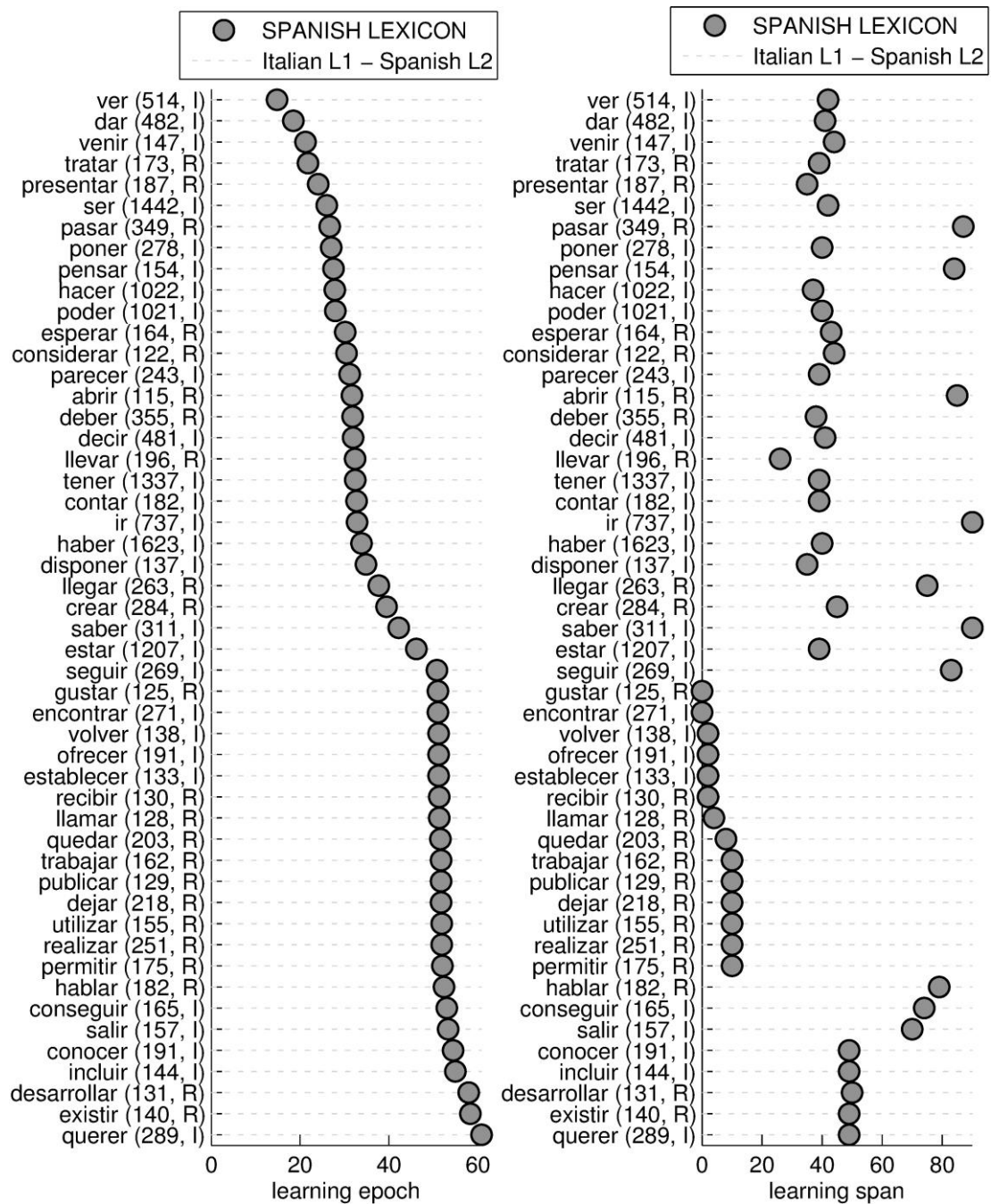
Figure 25: Spanish course (left plot) and span (right plot) of acquisition and in the Spanish_L2 and Italian_L1 condition, with paradigms ranked by increasing learning epoch (from top to bottom).

77
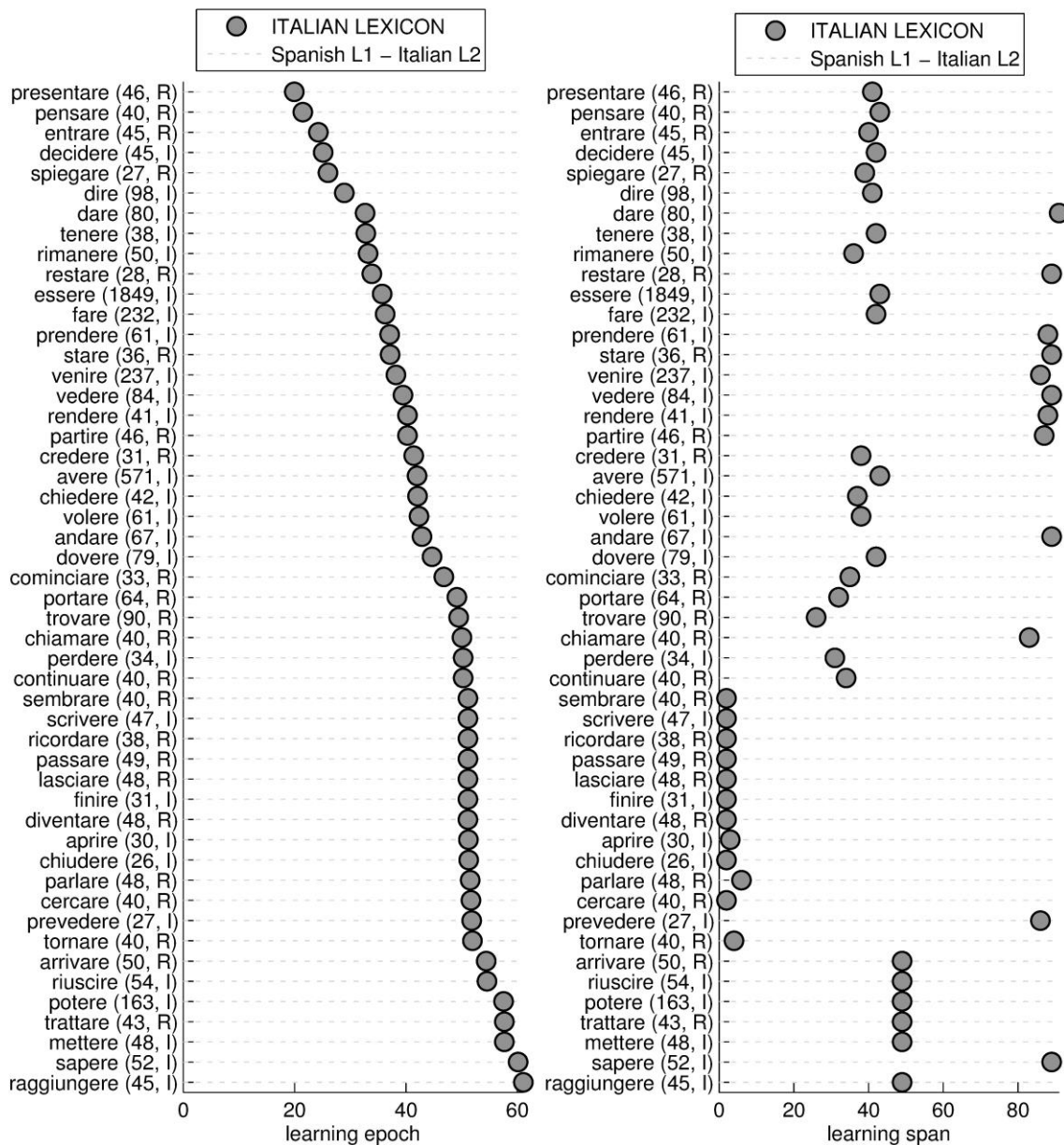
Figure 26: Italian course (left plot) and span (right plot) of acquisition and in the Spanish_L1 and Italian_L2 condition, with paradigms ranked by increasing learning epoch (from top to bottom).
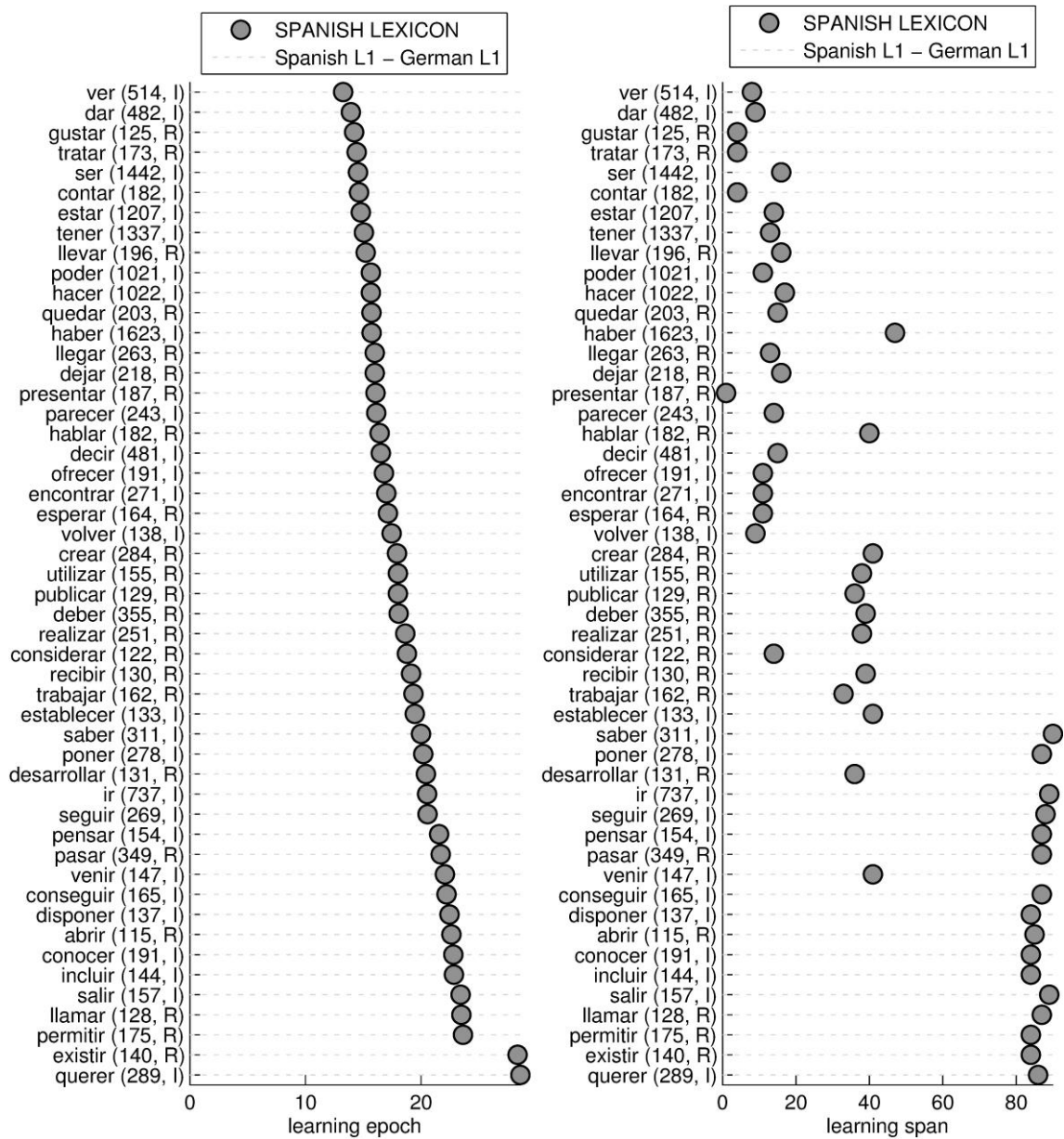
Figure 27: Spanish course (left plot) and span (right plot) of acquisition and in the Spanish_L1 and German_L1 condition, with paradigms ranked by increasing learning epoch (from top to bottom).
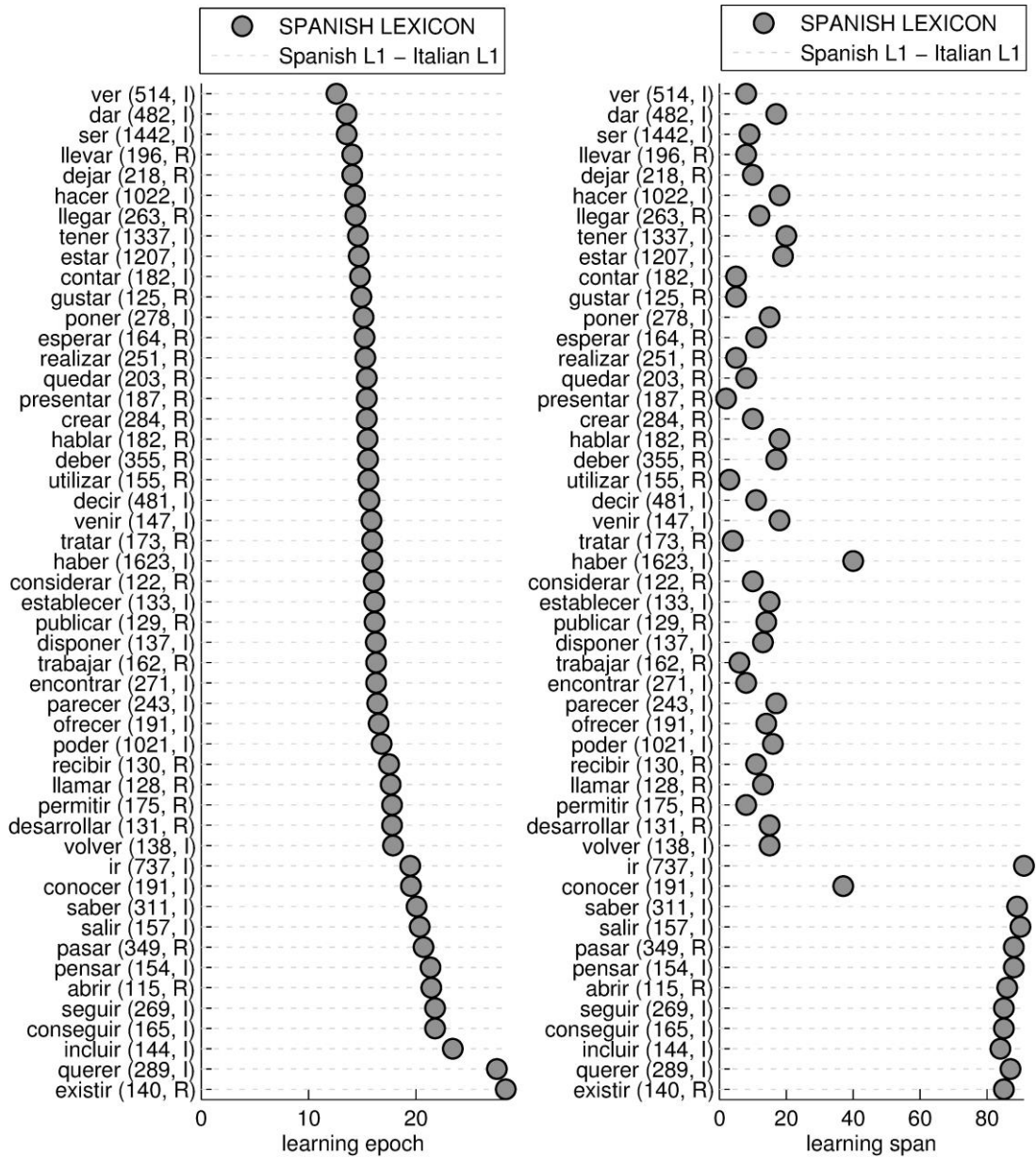
Figure 28: Spanish course (left plot) and span (right plot) of acquisition and in the Spanish_L1 and Italian_L1 condition, with paradigms ranked by increasing learning epoch (from top to bottom).

6. CONCLUDING REMARKS

Starting from the evidence provided by researchers at ComPhys Lab of the Institute for Computational Linguistics, Italian National Research Council (Pisa, ILC-CNR), the main goal of my thesis was to extend the application of computational modelling of language acquisition in monolingual and bilingual contexts to Spanish, which has not yet been treated within the given research framework.

For the first step, I briefly outlined some of the most prominent psycholinguistic approaches to the study of language acquisition. Secondly, three major models of morphological processing have been presented. For instance, three models of lexical representation and processing have been explained, following the classification proposed by Bybee (1995), i.e. dual-processing model, connectionist model, and network model. The difference between these three models lies in whether they make a distinction between regular and irregular verbs and their processing models, and whether or not the type/token frequency of verbal morphological patterns plays any role at all.

The experimental part of this study was focused on the first and second language acquisition of Spanish verbs, contrasted with parallel datasets in the Italian and German languages. In order to compile the dataset, I extracted the 50 most frequent verb paradigms from European Spanish Web Corpus (2011), available in Sketch Engine, for a total of 750 inflected forms (corresponding to the forms of the infinitive, present, and past participle, singular and plural simple present, singular and plural simple past). The frequency distribution is provided for each inflected form. For an analysis and evaluation of the emergent organization of paradigmatic relations, I annotated each form with morpho-syntactic information (i.e. stem and affix length, paradigmatic cell, formal (ir)regularity, paradigm). Specific difficulties arose during the segmentation of Spanish verbs, due to the peculiarities of some irregular patterns.

The computational modelling and processing of Spanish verbs forms has been simulated with Temporal Self-Organizing Maps (TSOMs), based on Kohonen's Self-Organizing Maps (2001), augmented with a temporal layer. Basically, this computational model reproduces dynamics of lexical learning and processing by imitating the emergence of neural self-organization, through the incremental adaptation of topologically and temporally aligned synaptic connections.

Starting from the literature review, also in connection with psycholinguistic evidence made available by studies of the last thirty years, I tried to put in evidence here that an adaptive self-organization during learning is conducive to the emergence of relations between word forms, which are stored in the mental lexicon in a concurrent and competitive dynamic. In particular, in the adopted bilingual perspective, monitoring the acquisitional trajectories of more than one lexica (in both L1+L2 and L1/L1 contexts) showed how recycled memory resources and weaker connections affect L2 acquisition and processing, with a smaller specialization for context-specific input chunks, depending on the exposure conditions.

With this goal in mind, experiments in different training conditions were designed. It is obvious that many other experimental conditions could be tested, as for example, more degrees of bilingualism, in a gradient of successive bilingualism ranging from very early, early, up to late, and very late bilingualism. This kind of approach would focus the attention on the plasticity loss and the increasing entrenchment of L1, with a subsequent and gradual difficulty for the L2 lexicon to create its own context-specific specialization.

Although time and space constraints did not allow these additional learning conditions, it is not hard to imagine and predict TSOMs behavior in these regimes. Map plasticity, as often underlined in my thesis, is in fact a determinant for a sufficiently specialized representation of lexical input. Very late exposure to L2 word forms, or early exposure to very reduced evidence of L2 lexicon, will determine a local, parasitic representation for the L2. As observed by Hernandez

et al. (2005), late bilinguals typically learn L2 with a reduced plasticity relying on strategies of "parasitic dependence" of L2 on their L1.

I may conclude that neurally-inspired computational models can provide a computational framework to analytically verify and study the developmental processes governing the acquisition and processing of the morphological lexicon in different languages, and reproduce a wide range of naturalistic conditions of both mono- and multi-lingual input exposure.

In such perspective, a more complex computational architecture, where a temporal self-organizing map is connected to other levels of representation, according to the Hebbian learning principle, may better simulate and predict speakers' and learners' lexical behavior. As evidenced in this study, the biologically-inspired TSOM architecture on the one hand provide a temporal layer that succeed in properly reproducing the sequential nature of linguistic inputs; on the other hand, however, it addresses this investigation from an exclusively phonotactic/orthotactic viewpoint, which neglects the semantic level in favor of the surface word relation analysis Therefore, my suggestion goes in the direction of a combination of the TSOM's level of representation of lexical input as temporal sequence of symbols with the DevLex II word's meaning representation.

Prospective research in this direction would give the possibility to simulate, and thus explain, more and more complex interaction and competition effects also at the level of semantic access in the frame of first and second language studies.

On a different dimension, a further development of such an approach could develop a more complex input representation, simulating not only single words but complex words (complex words, sentences).

REFERENCES

Aitchison, J. (1994). *Words in the Mind: An Introduction to the Mental Lexicon.* Oxford: Blackwell Publishers.

Albright, A. (2007). *Gradient phonological acceptability as a grammatical effect.* Camdridge, MS: MIT Press.
Retrieved from: http://web.mit.edu/albright/www/papers [Last access: 30/06/2017]

Albright, A. (2006). "Gradient phonotactic effects: lexical? grammatical? both? neither?" Paper presented at the 80th Annual LSA Meeting-Albuquerque, NM. Retrieved from: http://www.mit.edu/~albright/papers/Albright-GrammaticalGradience.pdf [Last access: 05.07.2017]

Anderson, B. (2014). *Computational Neuroscience and Cognitive Modelling. A Student's Introduction to Methods and Procedures.* London: SAGE Publishing.

Aronoff, M. (1994). *Morphology by itself: stems and inflectional classes.* Cambridge: MIT Press.

Baayen, R.H. (2007). "Storage and computation in the mental lexicon." In G. Jarema, & G. Libben (eds.), *The Mental Lexicon: Core Perspectives.* Amsterdam: Elsevier, 81–104.

Baayen, R.H., Feldman L.B. & Schreuder R. (2006). „Morphological influences on the recognition of monosyllabic monomorphemic words." *Journal of Memory and Language* 55/2, 290-313.

Baayen, R.H., Lieber R. & Schreuder R. (1997). "The morphological complexity of simplex nouns." *Linguistics* 35, 861-877.

Baayen, H.R., Piepenbrock R. & Gulikers L. (1995). *The CELEX Lexical Database* (CD-ROM). Philadelphia: Linguistic Data Consortium.

Baddeley A.D. (1986). *Working Memory.* Oxford: Oxford University Press.

Bailey, T.M. & Hahn U. (2001). "Determinants of wordlikeness: Phonotactics or lexical neighborhoods?" *Journal of Memory and Language* 44/4, 568-591.

Bates, E. & MacWhinney B. (1989). "Functionalism and the Competition Model." In B. MacWhinney & E. Bates (eds.), *The Crosslinguistic Study of Sentence Processing* New York: Cambridge University Press, 3-76.

Bates, E. & MacWhinney B. (1987). "Competition, variation, and language learning." In B. MacWhinney (eds.), *Mechanisms of language acquisition*, London, New York: Routledge, 157-193.

Bradlow, A.R. & Pisoni D.B. (1999). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors." *The Journal of the Acoustical Society of America* 106/4, 2074-2085.

Bybee, J. L. (2010). *Language, Usage and Cognition.* Cambridge: Cambridge University Press.

Bybee, J. L. (2001). *Phonology and Language Use.* Cambridge: Cambridge University Press.

Bybee, J. L. (1995). "Regular Morphology and the Lexicon." *Language and Cognitive Processes* 10, 425-455.

Bybee, J.L. (1988) "Morphology as lexical organization." In M. Hammond & M. Noonan (eds.), *Theoretical morphology.* San Diego: Academic Press, 119-141.

Bybee, J. (1985). *Morphology. A Study of the Relation between Meaning and Form.* Amsterdam: John Benjamins.

Caramazza, A., Laudanna, A. & Romani C. (1988). "Lexical access and inflectional morphology." *Cognition* 28/3, 297-332.

Chomsky, N. (1980). *Rules and Representati*ons. New York: Columbia University Press.

Chomsky, N. (1968). *Language and mind.* New York: Harcourt, Brace, Jovanovitch.

Christoffels, I. & De Groot, A. (2004). "Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing." *Bilingualism: Language and Cognition* 7/3, 227-240.

Cook, V. (2013). *Second language learning and language teaching.* New York: Routledge.

Cutler, A. (1980). "Productivity in word formation." In J. Kreiman, & A.E. Ojeda (eds.), *Papers from the Sixteenth Regional Meeting,* Chicago Linguistics Society, 45-51.

De Groot, A.M.B. & Van Hell, J.G. (2005). "The learning of foreign language vocabulary." In J.F. Kroll & A.M.B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches.* Oxford: Oxford University Press, 9-29.

Domínguez, A., Cuetos, F. & Segui, J. (2000). "Morphological processing in word recognition: A review with particular reference to Spanish data." *Psicológica* 21/2, 375-401.

Ellis, N.C. (2005). "Introduction to Part I." In J.F. Kroll & A.M.B. de Groot (eds.), *The Handbook of Bilingualism.* New York: Oxford University Press, 3-8.

Ellis, N.C. (2002). "Frequency effects in language processing." *Studies in Second Language Acquisition* 24, 143-188.

Ellis, N.C. (1998). "Emergentism, Connectionism and Language Learning." *Language Learning* 48/4, 631-664.

Ellis, N.C. & Schmidt R. (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, J.L. (2004). "An Alternative view of the Mental Lexicon." *Trends of Cognitive Science* 8/7, 301-306.

Elman, J.L. (1995). "Language as a dynamical system." In R.F. Port & T. Van Gelder (eds.) *Mind as motion: Explorations in the dynamics of cognition.* Cambridge, MA: MIT Press, 496-518.

Embick, D. (2012). "Contextual conditions on stem alternations: Illustrations from the Spanish conjugation." In I. Franco, S. Lusini & A. Saab (eds.) *Romance Languages and Linguistic Theory 2010: Selected papers from 'Going Romance' Leiden 2010.* Romance Languages and Linguistic Theory 4, 21–40.

Ferro M., Marzi C. & Pirrelli V. (2011). "A Self-Organizing Model of Word Storage and Processing: Implications for Morphology Learning." *Lingue e Linguaggio* 10/2, 209- 226.

Ferro, M., Pezzulo G. & Pirrelli V. (2010). "Morphology, memory, and the mental lexicon." *Lingue e Linguaggio* 9/2, 199-238.

Fodor, J.A. (1983). *The modularity of mind.* Cambridge, Ma: MIT Press.

Gagné, C. (2017). "Psycholinguistic Approaches to Morphology." Oxford Research Encyclopedia of Linguistics. Retrieved from: http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-258 [Last access: 03.06.2017]

Grainger, J., Colé P. & Segui J. (1991). "Masked morphological priming in visual word recognition." *Journal of Memory and Language* 30/3, 370-384.

Hebb, D. O. (1949) *Organization of behavior.* New York: Wiley.

Hernandez, A., Li P. & MacWhinney B. (2005). "The emergence of competing modules in bilingualism." *Trends in cognitive sciences* 9/5, 220-225.

Holmén, J. (1996). "Process modelling using Self-Organizing Map." Master's thesis. Retrieved from: http://users.ics.aalto.fi/jhollmen/dippa/dippa.html [Last access: 03.06.2017]

Jackendoff, R. (1975). "Morphological and semantic regularities in the lexicon." *Language* 51, 639-671.

Kaas J., Merzenich M. & Killackey H. (1983). "The reorganization of somatosensory cortex following peripheral nerve damage in adult and developing mammals." *Annual Review of Neuroscience* 6, 325-356.

Kohonen, T. (2001). *Self-organizing maps.* New York: Springer.

Kohonen, T. (1982). „Self-organized formation of topologically correct feature maps." *T. Biol. Cybern* 43, 59-69.

Kuperman, V., Schreuder R., Bertram R. & Baayen R.H. (2009). „Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing." *Journal of Experimental Psychology: Human Perception and Performance* 35, 876-895.

Li, P. (2009). "Lexical Organization and Competition in First and Second Languages: Computational and Neural Mechanism." *Cognitive Science* 33, 629-664.

Li, P., Zhao X. & MacWhinney B. (2007). "Dynamic SelfOrganization and Early Lexical Development in Children." *Cognitive science* 31/4, 581-612.

Libben, G. (2005). "Everything is psycholinguistics: Material and methodological considerations in the study of compound processing." *Canadian Journal of Linguistics* 50, 267-283.

Lohnas, L.J. & Kahana M.J. (2013). "Parametric effects of word frequency in memory for mixed frequency lists." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39, 1943-1946.

Luce, P.A. (1986). "Neighborhoods of words in the mental lexicon." Doctoral dissertation, Indian University, Bloomington.

Lyding, V., Stemle, E., Borghetti, C., Brunello, M., Castagnoli, S., Dell'Orletta, F., Dittmann, H., Lenci, A. & Pirrelli, V. (2014). "The PAISÀ Corpus of Italian Web Texts." In F. Bildhauer & R. Schäfer (eds.), *Proceedings of the 9th Web as Corpus Workshop* (WaC-9). Gothenburg, 36-43.

MacWhinney, B. (2012). "The Logic of the Unified Model." In Handbook of Second Language Acquisition. London: Routledge, 211-227.

MacWhinney, B. (2005). "A Unified Model of Language Acquisition." *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press, 49-67.

MacWhinney B. (2004). "A unified model of language acquisition." In Kroll J. & A. De Groot (eds.) *Handbook of bilingualism: Psycholinguistic approaches.* Oxford: Oxford University Press.

MacWhinney, B. (2001). "The competition model: The input, the context, and the brain." In P. Robinson (eds.) *Cognition and Second Language Instruction.* Cambridge: Cambridge University Press, 69-90.

MacWhinney, B. (1999) *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (1992) "The Competition Model and Foreign Language Acquisition." In M.L. Swartz & M. Yazdani (eds.) *Intelligent Tutoring Systems for Foreign Language Learning*. NATO ASI Series (Series F: Computer and Systems Sciences), vol 80. Berlin, Heidelberg: Springer.

Mandler, G., Goodman G.O. & Wilkes-Gibbs D.L. (1982). "The word-frequency paradox in recognition." *Memory & Cognition* 10, 33-42.

Marcus, G.F., Ullman M., Pinker S., Hollander M. & Rosen T.J. (1990). *Overregularization*. Cambridge, MA: MIT. Occasional paper n. 41.

Marslen-Wilson W. (1993). "Issues of representation in lexical access." *Cognitive models of speech processing: The second Sperlonga meeting*, 187-210.

Marzi, C., Ferro M. & Nahli O. (2017) "Arabic word processing and morphology induction through adaptive memory self-organisation strategies." *Journal of King Saud University. Computer and information sciences* 29/2, 179-188. Elsevier. Retrieved from: http://www.sciencedirect.com/science/article/pii/S1319157816301148 [Last access: 14.06.2017]

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

Marzi, C. & Pirrelli V. (2015). "A Neuro-Computational Approach to Understanding the Mental Lexicon." *Journal of Cognitive Science* 16/4, 493-534.

Marzi, C., Ferro M. & Pirrelli V. (2014a). "Morphological structure through lexical parsability." *Lingue e Linguaggio* 13/2, 263-290.

Marzi, C., Ferro M. & Keuleers E. (2014b). "Perception of typicality in the lexicon: wordlikeness, lexical density and morphonotactic constraints." *Suvremena lingvistika* 40/78, 171-191.

Marzi, C., Ferro M. & Pirrelli V. (2012). "Word alignment and paradigm induction." In C. Marzi & V. Pirrelli (eds.) *Lingue e Linguaggio* 11/2. Bologna: il Mulino, 251-274.

Matthews, P.H. (1993). *Grammatical Theory in the United States: from Bloomfield to Chomsky*. Cambridge: Cambridge University Press.

McClelland, J. L., Rumelhart, D. E., & the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Volume II. Cambridge, MA: MIT Press, 216-271.

Milin, P., Đurđević, D.F. & Moscoso del Prado Martín, F. (2009). "The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian." *Journal of Memory and Language* 60/1, 50-64.

Murre, J.M.J. (2005). "Models of monolingual and bilingual language acquisition." In Kroll J.F. & A.M.B. De Groot (eds.) *Handbook of bilingualism: Psycholinguistic approaches.* New York: Oxford University Press, 154-169.

Nosofsky, R.M. (1986). "Attention, similarity, and the identification–categorization relationship." *Journal of experimental psychology: General* 115/1, 39.

Nosofsky, R.M. (1990). "Relations between exemplar-similarity and likelihood models of classification." *Journal of Mathematical Psychology* 34/4, 393-418.

Real Academia Española. (2010) *Nueva gramática de la lengua española.* Madrid: Espasa Libros.

Pinker, S. & Prince A. (1991). "Regular and Irregular Morphology and the Psychological Status of Rules of Grammar." *Annual Meeting of the Berkeley Linguistics Society* 17/1, 230-251.

Pirrelli V., Ferro M. & Marzi C. (2015). "Computational complexity of abstractive morphology." In M. Baerman, D. Brown & G. Corbett (eds.), *Understanding and Measuring Morphological Complexity.* Oxford: Oxford University Press, 141-166.

Pirrelli V., Marzi C. & Ferro M. (2014). "Two-dimensional Wordlikeness Effects in Lexical Organisation." In R. Basili, A. Lenci & B. Magnini (eds.) *Proceedings of the First Italian Conference on Computational Linguistic.* Pisa University Press, 301-305.

Plaut, D.C. & Gonnerman L.M. (2000). "Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing?" *Language and Cognitive Processes* 15/4/5, 445-485.

Pulvermüller F. (2002). *The Neuroscience of Language. On brain circuits of Words and Serial Order*. Cambridge: Cambridge University Press.

Rastle, K. & Davis M.H. (2008). "Morphological decomposition based on the analysis of orthography." *Language and Cognitive Processes* 23/7/8, 942-971.

Rojas, R. (1996). *Neural Networks. A systematic introduction*. Springer-Verlag, Berlin. 392-412.

Rumelhart, D. & McClelland J. (1987). "Learning the past tenses of English verbs: Implicit rules or parallel distributed processing?" In MacWhinney B. (eds.) *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum, 195-248.

Rumelhart D. & McClelland J. (1986). "On learning the past tense of English verbs." In D.E. Rumelhart & J. McClelland (eds.) *Parallel distributed processing: Explanations in the microstructure of cognition*. Cambridge: MIT Press, 216-271.

Sandra D. (1994). "The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective*." Language and Cognitive Processes* 9/3, 227-269.

Schreuder R. & Baayen H. (1997). "How complex simplex words can be." *Journal of Memory and Language* 37, 118-139.

Segalowitz, N. & Hulstijn, J. (2005). "Automaticity in bilingualism and second language learning." In J.F. Kroll & A.M.B. De Groot, (eds.) *Handbook of bilingualism: Psycholinguistic approaches*. Oxford, UK: Oxford University Press, 371-388.

Snow, C.E. (1999). "Social perspectives on the emergence of language." In B. MacWhinney (eds.) *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum Associates, 257-276.

Stemberger, J.P. (1989). "Speech errors in early child language production." *Journal of Memory and Language* 28/2, 164-188.

Taft, M. & Forster K. (1976). "Lexical storage and retrieval of polymorphemic and polysyllabic words." *Journal of Verbal Learning and Verbal Behavior* 15, 607-620.

Thomas, N.J.T. (2017) "Mental Imagery." *The Stanford Encyclopedia of Philosophy*, E.N. Zalta (ed.). Retrieved from: https://plato.stanford.edu/archives/spr2017/entries/mental-imagery [Last access: 29.06.2017]

Traxler, M.J. (2012*). Introduction to psycholinguistics: Understanding language science.* Chichester, West Sussex, UK; Malden, MA: Wiley-Blackwell.

Välimaa-Blum, R. (2009). "Phonotactic constraints in cognitive phonology." Corela, 7-1. Retrieved from: http://corela.revues.org/137 [Last access: 05.07.2017]

Van Heuven, W.J.B., Dijkstra T. & Grainger J. (1998). "Orthographic Neighborhood Effects in Bilingual Word Recognition." *Journal of Memory and Language* 39, 458-483.

Vennemann, T. (1974). "Words and syllables in natural generative grammar." In A. Bruck, R.A. Fox & M.W. La Galy (eds.) *Papers from the parasession on Natural Phonology*. Chicago: Chicago Linguistic Society, 346-374.

Waskan, J. (2010). "Connectionism." Internet Encyclopedia of Philosophy. Retrieved from: http://www.iep.utm.edu/connect/ [Last access: 1/07/2017]

Zamorano-Mansilla, J.R. (2008). "Dealing with verbal inflection in natural language generation: Resources for minimal requirement." *Revista Española de Lingüística Aplicada* 21, 289-314.

Zhao, X. & Li P. (2007). "Bilingual Lexical Representation in a Self-Organizing Neural Network Model." *Proceedings of the Cognitive Science Society* 29.

ANNEX 1

SYMBOL CODING

#	1000000000000000000000000000000000
A	0100000000000000000000000000000000
B	0010000000000000000000000000000000
C	0001000000000000000000000000000000
D	0000100000000000000000000000000000
E	0000010000000000000000000000000000
F	0000001000000000000000000000000000
G	0000000100000000000000000000000000
H	0000000010000000000000000000000000
I	0000000001000000000000000000000000
J	0000000000100000000000000000000000
K	0000000000010000000000000000000000
L	0000000000001000000000000000000000
M	0000000000000100000000000000000000
N	0000000000000010000000000000000000
O	0000000000000001000000000000000000
P	0000000000000000100000000000000000
Q	0000000000000000010000000000000000
R	0000000000000000001000000000000000
S	0000000000000000000100000000000000
T	0000000000000000000010000000000000
U	0000000000000000000001000000000000
V	0000000000000000000000100000000000
W	0000000000000000000000010000000000
X	0000000000000000000000001000000000
Y	0000000000000000000000000100000000
Z	0000000000000000000000000010000000

$	0000000000000000000000000100000000
'	0000000000000000000000000010000000
ae	0000000000000000000000000001000000
oe	0000000000000000000000000000100000
ue	0000000000000000000000000000010000
ss	0000000000000000000000000000001000
í	0000000000000000000000000000000100
ó	0000000000000000000000000000000010
á	0000000000000000000000000000000001

ANNEX 2

SPANISH DATASET

tsom word form	frequency	word form	paradigmatic cell	conjugation	lexical exponent	prefix length	root length	suffix length	I/R
#,A,B,R,I,R,$	22	abrir	i	1	abrir	0	3	2	R
#,A,B,R,I,E,N,D,O,$	5	abriendo	pE	1	abrir	0	3	5	R
#,A,B,I,E,R,T,O,$	36	abierto	pA	1	abrir	0	5	2	R
#,A,B,R,O,$	2	abro	1SIE	1	abrir	0	3	1	R
#,A,B,R,E,S,$	2	abres	2SIE	1	abrir	0	3	2	R
#,A,B,R,E,$	18	abre	3SIE	1	abrir	0	3	1	R
#,A,B,R,I,M,O,S,$	3	abrimos	1PIE	1	abrir	0	3	4	R
#,A,B,R,í,S,$	2	abrís	2PIE	1	abrir	0	3	2	R
#,A,B,R,E,N,$	6	abren	3PIE	1	abrir	0	3	2	R
#,A,B,R,í,$	2	abrí	1SIA	1	abrir	0	3	1	R
#,A,B,R,I,S,T,E,$	2	abriste	2SIA	1	abrir	0	3	4	R
#,A,B,R,I,ó,$	9	abrió	3SIA	1	abrir	0	3	2	R
#,A,B,R,I,M,O,S,$	1	abrimos	1PIA	1	abrir	0	3	4	R
#,A,B,R,I,S,T,E,I,S,$	2	abristeis	2PIA	1	abrir	0	3	6	R
#,A,B,R,I,E,R,O,N,$	3	abrieron	3PIA	1	abrir	0	3	5	R
#,C,O,N,O,C,E,R,$	80	conocer	i	2	conocer	0	5	2	I
#,C,O,N,O,C,I,E,N,D,O,$	5	conociendo	pE	2	conocer	0	5	5	I
#,C,O,N,O,C,I,D,O,$	33	conocido	pA	2	conocer	0	5	3	I
#,C,O,N,O,Z,C,O,$	3	conozco	1SIE	2	conocer	0	6	1	I
#,C,O,N,O,C,E,S,$	5	conoces	2SIE	2	conocer	0	5	2	I
#,C,O,N,O,C,E,$	24	conoce	3SIE	2	conocer	0	5	1	I
#,C,O,N,O,C,E,M,O,S,$	10	conocemos	1PIE	2	conocer	0	5	4	I
#,C,O,N,O,C,E,',I,S,$	2	conocéis	2PIE	2	conocer	0	5	4	I
#,C,O,N,O,C,E,N,$	11	conocen	3PIE	2	conocer	0	5	2	I
#,C,O,N,O,C,í,$	4	conocí	1SIA	2	conocer	0	5	1	I
#,C,O,N,O,C,I,S,T,E,$	2	conociste	2SIA	2	conocer	0	5	4	I
#,C,O,N,O,C,I,ó,$	5	conoció	3SIA	2	conocer	0	5	2	I
#,C,O,N,O,C,I,M,O,S,$	2	conocimos	1PIA	2	conocer	0	5	4	I
#,C,O,N,O,C,I,S,T,E,I,S,$	2	conocisteis	2PIA	2	conocer	0	5	6	I
#,C,O,N,O,C,I,E,R,O,N,$	3	conocieron	3PIA	2	conocer	0	5	5	I
#,C,O,N,S,E,G,U,I,R,$	59	conseguir	i	3	conseguir	0	7	2	I
#,C,O,N,S,I,G,U,I,E,N,D,O,$	7	consiguiendo	pE	3	conseguir	0	7	5	I
#,C,O,N,S,E,G,U,I,D,O,$	28	conseguido	pA	3	conseguir	0	7	3	I
#,C,O,N,S,I,G,O,$	8	consigo	1SIE	3	conseguir	0	6	1	I
#,C,O,N,S,I,G,U,E,S,$	3	consigues	2SIE	3	conseguir	0	7	2	I
#,C,O,N,S,I,G,U,E,$	18	consigue	3SIE	3	conseguir	0	7	1	I
#,C,O,N,S,E,G,U,I,M,O,S,$	2	conseguimos	1PIE	3	conseguir	0	7	4	I
#,C,O,N,S,E,G,U,í,S,$	2	conseguís	2PIE	3	conseguir	0	7	2	I
#,C,O,N,S,I,G,U,E,N,$	7	consiguen	3PIE	3	conseguir	0	7	2	I
#,C,O,N,S,E,G,U,í,$	3	conseguí	1SIA	3	conseguir	0	7	1	I
#,C,O,N,S,E,G,U,I,S,T,E,$	2	conseguiste	2SIA	3	conseguir	0	7	4	I

#,C,O,N,S,I,G,U,I,ó,$	14	consiguió	3SIA	3	conseguir	0	7	2	I
#,C,O,N,S,E,G,U,I,M,O,S,$	4	conseguimos	1PIA	3	conseguir	0	7	4	I
#,C,O,N,S,E,G,U,I,S,T,E,I,S,$	2	conseguisteis	2PIA	3	conseguir	0	7	6	I
#,C,O,N,S,I,G,U,I,E,R,O,N,$	6	consiguieron	3PIA	3	conseguir	0	7	5	I
#,C,O,N,S,I,D,E,R,A,R,$	16	considerar	i	4	considerar	0	8	2	R
#,C,O,N,S,I,D,E,R,A,N,D,O,$	12	considerando	pE	4	considerar	0	8	4	R
#,C,O,N,S,I,D,E,R,A,D,O,$	16	considerado	pA	4	considerar	0	8	3	R
#,C,O,N,S,I,D,E,R,O,$	9	considero	1SIE	4	considerar	0	8	1	R
#,C,O,N,S,I,D,E,R,A,S,$	2	consideras	2SIE	4	considerar	0	8	2	R
#,C,O,N,S,I,D,E,R,A,$	33	considera	3SIE	4	considerar	0	8	1	R
#,C,O,N,S,I,D,E,R,A,M,O,S,$	7	consideramos	1PIE	4	considerar	0	8	4	R
#,C,O,N,S,I,D,E,R,á,I,S,$	2	consideráis	2PIE	4	considerar	0	8	3	R
#,C,O,N,S,I,D,E,R,A,N,$	14	consideran	3PIE	4	considerar	0	8	2	R
#,C,O,N,S,I,D,E,R,E,',$	2	consideré	1SIA	4	considerar	0	8	2	R
#,C,O,N,S,I,D,E,R,A,S,T,E,$	2	consideraste	2SIA	4	considerar	0	8	4	R
#,C,O,N,S,I,D,E,R,ó,$	2	consideró	3SIA	4	considerar	0	8	1	R
#,C,O,N,S,I,D,E,R,A,M,O,S,$	1	consideramos	1PIA	4	considerar	0	8	4	R
#,C,O,N,S,I,D,E,R,A,S,T,E,I,S,$	2	considerasteis	2PIA	4	considerar	0	8	6	R
#,C,O,N,S,I,D,E,R,A,R,O,N,$	2	consideraron	3PIA	4	considerar	0	8	4	R
#,C,O,N,T,A,R,$	36	contar	i	5	contar	0	4	2	I
#,C,O,N,T,A,N,D,O,$	9	contando	pE	5	contar	0	4	4	I
#,C,O,N,T,A,D,O,$	11	contado	pA	5	contar	0	4	3	I
#,C,U,E,N,T,O,$	7	cuento	1SIE	5	contar	0	5	1	I
#,C,U,E,N,T,A,S,$	4	cuentas	2SIE	5	contar	0	5	2	I
#,C,U,E,N,T,A,$	63	cuenta	3SIE	5	contar	0	5	1	I
#,C,O,N,T,A,M,O,S,$	9	contamos	1PIE	5	contar	0	4	4	I
#,C,O,N,T,á,I,S,$	2	contáis	2PIE	5	contar	0	4	3	I
#,C,U,E,N,T,A,N,$	21	cuentan	3PIE	5	contar	0	5	2	I
#,C,O,N,T,E,',$	2	conté	1SIA	5	contar	0	4	2	I
#,C,O,N,T,A,S,T,E,$	2	contaste	2SIA	5	contar	0	4	4	I
#,C,O,N,T,ó,$	10	contó	3SIA	5	contar	0	4	1	I
#,C,O,N,T,A,M,O,S,$	1	contamos	1PIA	5	contar	0	4	4	I
#,C,O,N,T,A,S,T,E,I,S,$	2	contasteis	2PIA	5	contar	0	4	6	I
#,C,O,N,T,A,R,O,N,$	3	contaron	3PIA	5	contar	0	4	4	I
#,C,R,E,A,R,$	54	crear	i	6	crear	0	3	2	R
#,C,R,E,A,N,D,O,$	11	creando	pE	6	crear	0	3	4	R
#,C,R,E,A,D,O,$	4	creado	pA	6	crear	0	3	3	R
#,C,R,E,O,$	162	creo	1SIE	6	crear	0	3	1	R
#,C,R,E,A,S,$	3	creas	2SIE	6	crear	0	3	2	R
#,C,R,E,A,$	17	crea	3SIE	6	crear	0	3	1	R
#,C,R,E,A,M,O,S,$	3	creamos	1PIE	6	crear	0	3	4	R
#,C,R,E,á,I,S,$	2	creáis	2PIE	6	crear	0	3	3	R
#,C,R,E,A,N,$	8	crean	3PIE	6	crear	0	3	2	R
#,C,R,E,E,',$	2	creé	1SIA	6	crear	0	3	2	R
#,C,R,E,A,S,T,E,$	2	creaste	2SIA	6	crear	0	3	4	R
#,C,R,E,ó,$	9	creó	3SIA	6	crear	0	3	1	R
#,C,R,E,A,M,O,S,$	1	creamos	1PIA	6	crear	0	3	4	R
#,C,R,E,A,S,T,E,I,S,$	2	creasteis	2PIA	6	crear	0	3	6	R

#,C,R,E,A,R,O,N,$	4	crearon	3PIA	6	crear	0	3	4	R
#,D,A,R,$	116	dar	i	7	dar	0	1	2	I
#,D,A,N,D,O,$	30	dando	pE	7	dar	0	1	4	I
#,D,A,D,O,$	75	dado	pA	7	dar	0	1	3	I
#,D,O,Y,$	12	doy	1SIE	7	dar	0	1	2	I
#,D,A,S,$	9	das	2SIE	7	dar	0	1	2	I
#,D,A,$	118	da	3SIE	7	dar	0	1	1	I
#,D,A,M,O,S,$	9	damos	1PIE	7	dar	0	1	4	I
#,D,A,I,S,$	3	dais	2PIE	7	dar	0	1	3	I
#,D,A,N,$	42	dan	3PIE	7	dar	0	1	2	I
#,D,I,$	11	di	1SIA	7	dar	0	1	1	I
#,D,I,S,T,E,$	2	diste	2SIA	7	dar	0	1	4	I
#,D,I,O,$	34	dio	3SIA	7	dar	0	1	2	I
#,D,I,M,O,S,$	3	dimos	1PIA	7	dar	0	1	4	I
#,D,I,S,T,E,I,S,$	2	disteis	2PIA	7	dar	0	1	6	I
#,D,I,E,R,O,N,$	16	dieron	3PIA	7	dar	0	1	5	I
#,D,E,B,E,R,$	2	deber	i	8	deber	0	3	2	R
#,D,E,B,I,E,N,D,O,$	6	debiendo	pE	8	deber	0	3	5	R
#,D,E,B,I,D,O,$	55	debido	pA	8	deber	0	3	3	R
#,D,E,B,O,$	11	debo	1SIE	8	deber	0	3	1	R
#,D,E,B,E,S,$	13	debes	2SIE	8	deber	0	3	2	R
#,D,E,B,E,$	148	debe	3SIE	8	deber	0	3	1	R
#,D,E,B,E,M,O,S,$	30	debemos	1PIE	8	deber	0	3	4	R
#,D,E,B,E,',I,S,$	3	debéis	2PIE	8	deber	0	3	4	R
#,D,E,B,E,N,$	69	deben	3PIE	8	deber	0	3	2	R
#,D,E,B,í,$	2	debí	1SIA	8	deber	0	3	1	R
#,D,E,B,I,S,T,E,$	2	debiste	2SIA	8	deber	0	3	4	R
#,D,E,B,I,ó,$	7	debió	3SIA	8	deber	0	3	2	R
#,D,E,B,I,M,O,S,$	2	debimos	1PIA	8	deber	0	3	4	R
#,D,E,B,I,S,T,E,I,S,$	2	debisteis	2PIA	8	deber	0	3	6	R
#,D,E,B,I,E,R,O,N,$	3	debieron	3PIA	8	deber	0	3	5	R
#,D,E,C,I,R,$	176	decir	i	9	decir	0	3	2	I
#,D,I,C,I,E,N,D,O,$	18	diciendo	pE	9	decir	0	3	5	I
#,D,I,C,H,O,$	13	dicho	pA	9	decir	0	4	1	I
#,D,I,G,O,$	6	digo	1SIE	9	decir	0	3	1	I
#,D,I,C,E,S,$	16	dices	2SIE	9	decir	0	3	2	I
#,D,I,C,E,$	103	dice	3SIE	9	decir	0	3	1	I
#,D,E,C,I,M,O,S,$	6	decimos	1PIE	9	decir	0	3	4	I
#,D,E,C,í,S,$	3	decís	2PIE	9	decir	0	3	2	I
#,D,I,C,E,N,$	38	dicen	3PIE	9	decir	0	3	2	I
#,D,I,J,E,$	13	dije	1SIA	9	decir	0	3	1	I
#,D,I,J,I,S,T,E,$	2	dijiste	2SIA	9	decir	0	3	4	I
#,D,I,J,O,$	73	dijo	3SIA	9	decir	0	3	1	I
#,D,I,J,I,M,O,S,$	3	dijimos	1PIA	9	decir	0	3	4	I
#,D,I,J,I,S,T,E,I,S,$	2	dijisteis	2PIA	9	decir	0	3	6	I
#,D,I,J,E,R,O,N,$	9	dijeron	3PIA	9	decir	0	3	4	I
#,D,E,J,A,R,$	54	dejar	i	10	dejar	0	3	2	R
#,D,E,J,A,N,D,O,$	17	dejando	pE	10	dejar	0	3	4	R

#,D,E,J,A,D,O,$	26	dejado	pA	10	dejar	0	3	3	R
#,D,E,J,O,$	12	dejo	1SIE	10	dejar	0	3	1	R
#,D,E,J,A,S,$	3	dejas	2SIE	10	dejar	0	3	2	R
#,D,E,J,A,$	42	deja	3SIE	10	dejar	0	3	1	R
#,D,E,J,A,M,O,S,$	9	dejamos	1PIE	10	dejar	0	3	4	R
#,D,E,J,á,I,S,$	2	dejáis	2PIE	10	dejar	0	3	3	R
#,D,E,J,A,N,$	16	dejan	3PIE	10	dejar	0	3	2	R
#,D,E,J,E,',$	4	dejé	1SIA	10	dejar	0	3	2	R
#,D,E,J,A,S,T,E,$	2	dejaste	2SIA	10	dejar	0	3	4	R
#,D,E,J,ó,$	20	dejó	3SIA	10	dejar	0	3	1	R
#,D,E,J,A,M,O,S,$	1	dejamos	1PIA	10	dejar	0	3	4	R
#,D,E,J,A,S,T,E,I,S,$	2	dejasteis	2PIA	10	dejar	0	3	6	R
#,D,E,J,A,R,O,N,$	8	dejaron	3PIA	10	dejar	0	3	4	R
#,D,E,S,A,R,R,O,L,L,A,R,$	40	desarrollar	i	11	desarrollar	0	9	2	R
#,D,E,S,A,R,R,O,L,L,A,N,D,O,$	12	desarrollando	pE	11	desarrollar	0	9	4	R
#,D,E,S,A,R,R,O,L,L,A,D,O,$	24	desarrollado	pA	11	desarrollar	0	9	3	R
#,D,E,S,A,R,R,O,L,L,O,$	2	desarrollo	1SIE	11	desarrollar	0	9	1	R
#,D,E,S,A,R,R,O,L,L,A,S,$	2	desarrollas	2SIE	11	desarrollar	0	9	2	R
#,D,E,S,A,R,R,O,L,L,A,$	20	desarrolla	3SIE	11	desarrollar	0	9	1	R
#,D,E,S,A,R,R,O,L,L,A,M,O,S,$	2	desarrollamos	1PIE	11	desarrollar	0	9	4	R
#,D,E,S,A,R,R,O,L,L,á,I,S,$	2	desarrolláis	2PIE	11	desarrollar	0	9	3	R
#,D,E,S,A,R,R,O,L,L,A,N,$	11	desarrollan	3PIE	11	desarrollar	0	9	2	R
#,D,E,S,A,R,R,O,L,L,E,',$	2	desarrollé	1SIA	11	desarrollar	0	9	2	R
#,D,E,S,A,R,R,O,L,L,A,S,T,E,$	2	desarrollaste	2SIA	11	desarrollar	0	9	4	R
#,D,E,S,A,R,R,O,L,L,ó,$	6	desarrolló	3SIA	11	desarrollar	0	9	1	R
#,D,E,S,A,R,R,O,L,L,A,M,O,S,$	1	desarrollamos	1PIA	11	desarrollar	0	9	4	R
#,D,E,S,A,R,R,O,L,L,A,S,T,E,I,S,$	2	desarrollasteis	2PIA	11	desarrollar	0	9	6	R
#,D,E,S,A,R,R,O,L,L,A,R,O,N,$	3	desarrollaron	3PIA	11	desarrollar	0	9	4	R
#,D,I,S,P,O,N,E,R,$	17	disponer	i	12	disponer	0	6	2	I
#,D,I,S,P,O,N,I,E,N,D,O,$	3	disponiendo	pE	12	disponer	0	6	5	I
#,D,I,S,P,U,E,S,T,O,$	33	dispuesto	pA	12	disponer	0	7	2	I
#,D,I,S,P,O,N,G,O,$	4	dispongo	1SIE	12	disponer	0	7	1	I
#,D,I,S,P,O,N,E,S,$	2	dispones	2SIE	12	disponer	0	6	2	I
#,D,I,S,P,O,N,E,$	44	dispone	3SIE	12	disponer	0	6	1	I
#,D,I,S,P,O,N,E,M,O,S,$	7	disponemos	1PIE	12	disponer	0	6	4	I
#,D,I,S,P,O,N,E,',I,S,$	2	disponéis	2PIE	12	disponer	0	6	4	I
#,D,I,S,P,O,N,E,N,$	12	disponen	3PIE	12	disponer	0	6	2	I
#,D,I,S,P,U,S,E,$	2	dispuse	1SIA	12	disponer	0	6	1	I
#,D,I,S,P,U,S,I,S,T,E,$	2	dispusiste	2SIA	12	disponer	0	6	4	I
#,D,I,S,P,U,S,O,$	3	dispuso	3SIA	12	disponer	0	6	1	I
#,D,I,S,P,U,S,I,M,O,S,$	2	dispusimos	1PIA	12	disponer	0	6	4	I
#,D,I,S,P,U,S,I,S,T,E,I,S,$	2	dispusisteis	2PIA	12	disponer	0	6	6	I
#,D,I,S,P,U,S,I,E,R,O,N,$	2	dispusieron	3PIA	12	disponer	0	6	5	I
#,E,N,C,O,N,T,R,A,R,$	60	encontrar	i	13	encontrar	0	7	2	I
#,E,N,C,O,N,T,R,A,N,D,O,$	3	encontrando	pE	13	encontrar	0	7	4	I
#,E,N,C,O,N,T,R,A,D,O,$	21	encontrado	pA	13	encontrar	0	7	3	I
#,E,N,C,U,E,N,T,R,O,$	11	encuentro	1SIE	13	encontrar	0	8	1	I
#,E,N,C,U,E,N,T,R,A,S,$	6	encuentras	2SIE	13	encontrar	0	8	2	I

#,E,N,C,U,E,N,T,R,A,$	77	encuentra	3SIE	13	encontrar	0	8	1	I
#,E,N,C,O,N,T,R,A,M,O,S,$	23	encontramos	1PIE	13	encontrar	0	7	4	I
#,E,N,C,O,N,T,R,á,I,S,$	2	encontráis	2PIE	13	encontrar	0	7	3	I
#,E,N,C,U,E,N,T,R,A,N,$	42	encuentran	3PIE	13	encontrar	0	8	2	I
#,E,N,C,O,N,T,R,E,',$	6	encontré	1SIA	13	encontrar	0	7	2	I
#,E,N,C,O,N,T,R,A,S,T,E,$	2	encontraste	2SIA	13	encontrar	0	7	4	I
#,E,N,C,O,N,T,R,ó,$	9	encontró	3SIA	13	encontrar	0	7	1	I
#,E,N,C,O,N,T,R,A,M,O,S,$	1	encontramos	1PIA	13	encontrar	0	7	4	I
#,E,N,C,O,N,T,R,A,S,T,E,I,S,$	2	encontrasteis	2PIA	13	encontrar	0	7	6	I
#,E,N,C,O,N,T,R,A,R,O,N,$	6	encontraron	3PIA	13	encontrar	0	7	4	I
#,E,S,P,E,R,A,R,$	29	esperar	i	14	esperar	0	5	2	R
#,E,S,P,E,R,A,N,D,O,$	15	esperando	pE	14	esperar	0	5	4	R
#,E,S,P,E,R,A,D,O,$	8	esperado	pA	14	esperar	0	5	3	R
#,E,S,P,E,R,O,$	48	espero	1SIE	14	esperar	0	5	1	R
#,E,S,P,E,R,A,S,$	3	esperas	2SIE	14	esperar	0	5	2	R
#,E,S,P,E,R,A,$	26	espera	3SIE	14	esperar	0	5	1	R
#,E,S,P,E,R,A,M,O,S,$	13	esperamos	1PIE	14	esperar	0	5	4	R
#,E,S,P,E,R,á,I,S,$	2	esperáis	2PIE	14	esperar	0	5	3	R
#,E,S,P,E,R,A,N,$	9	esperan	3PIE	14	esperar	0	5	2	R
#,E,S,P,E,R,E,',$	2	esperé	1SIA	14	esperar	0	5	2	R
#,E,S,P,E,R,A,S,T,E,$	2	esperaste	2SIA	14	esperar	0	5	4	R
#,E,S,P,E,R,ó,$	2	esperó	3SIA	14	esperar	0	5	1	R
#,E,S,P,E,R,A,M,O,S,$	1	esperamos	1PIA	14	esperar	0	5	4	R
#,E,S,P,E,R,A,S,T,E,I,S,$	2	esperasteis	2PIA	14	esperar	0	5	6	R
#,E,S,P,E,R,A,R,O,N,$	2	esperaron	3PIA	14	esperar	0	5	4	R
#,E,S,T,A,B,L,E,C,E,R,$	29	establecer	i	15	establecer	0	8	2	I
#,E,S,T,A,B,L,E,C,I,E,N,D,O,$	5	estableciendo	pE	15	establecer	0	8	5	I
#,E,S,T,A,B,L,E,C,I,D,O,$	33	establecido	pA	15	establecer	0	8	3	I
#,E,S,T,A,B,L,E,Z,C,O,$	2	establezco	1SIE	15	establecer	0	9	1	I
#,E,S,T,A,B,L,E,C,E,S,$	2	estableces	2SIE	15	establecer	0	8	2	I
#,E,S,T,A,B,L,E,C,E,$	30	establece	3SIE	15	establecer	0	8	1	I
#,E,S,T,A,B,L,E,C,E,M,O,S,$	2	establecemos	1PIE	15	establecer	0	8	4	I
#,E,S,T,A,B,L,E,C,E,',I,S,$	2	establecéis	2PIE	15	establecer	0	8	4	I
#,E,S,T,A,B,L,E,C,E,N,$	12	establecen	3PIE	15	establecer	0	8	2	I
#,E,S,T,A,B,L,E,C,í,$	2	establecí	1SIA	15	establecer	0	8	1	I
#,E,S,T,A,B,L,E,C,I,S,T,E,$	2	estableciste	2SIA	15	establecer	0	8	4	I
#,E,S,T,A,B,L,E,C,I,ó,$	5	estableció	3SIA	15	establecer	0	8	2	I
#,E,S,T,A,B,L,E,C,I,M,O,S,$	2	establecimos	1PIA	15	establecer	0	8	4	I
#,E,S,T,A,B,L,E,C,I,S,T,E,I,S,$	2	establecisteis	2PIA	15	establecer	0	8	6	I
#,E,S,T,A,B,L,E,C,I,E,R,O,N,$	3	establecieron	3PIA	15	establecer	0	8	5	I
#,E,S,T,A,R,$	131	estar	i	16	estar	0	3	2	I
#,E,S,T,A,N,D,O,$	12	estando	pE	16	estar	0	3	4	I
#,E,S,T,A,D,O,$	40	estado	pA	16	estar	0	3	3	I
#,E,S,T,O,Y,$	99	estoy	1SIE	16	estar	0	3	2	I
#,E,S,T,á,S,$	22	estás	2SIE	16	estar	0	3	2	I
#,E,S,T,á,$	524	está	3SIE	16	estar	0	3	1	I
#,E,S,T,A,M,O,S,$	93	estamos	1PIE	16	estar	0	3	4	I
#,E,S,T,á,I,S,$	5	estáis	2PIE	16	estar	0	3	3	I

#,E,S,T,á,N,$	225	están	3PIE	16	estar	0	3	2	I
#,E,S,T,U,V,E,$	10	estuve	1SIA	16	estar	0	5	1	I
#,E,S,T,U,V,I,S,T,E,$	2	estuviste	2SIA	16	estar	0	5	4	I
#,E,S,T,U,V,O,$	28	estuvo	3SIA	16	estar	0	5	1	I
#,E,S,T,U,V,I,M,O,S,$	5	estuvimos	1PIA	16	estar	0	5	4	I
#,E,S,T,U,V,I,S,T,E,I,S,$	2	estuvisteis	2PIA	16	estar	0	5	6	I
#,E,S,T,U,V,I,E,R,O,N,$	9	estuvieron	3PIA	16	estar	0	5	5	I
#,E,X,I,S,T,I,R,$	10	existir	i	17	existir	0	5	2	R
#,E,X,I,S,T,I,E,N,D,O,$	4	existiendo	pE	17	existir	0	5	5	R
#,E,X,I,S,T,I,D,O,$	4	existido	pA	17	existir	0	5	3	R
#,E,X,I,S,T,O,$	2	existo	1SIE	17	existir	0	5	1	R
#,E,X,I,S,T,E,S,$	2	existes	2SIE	17	existir	0	5	2	R
#,E,X,I,S,T,E,$	58	existe	3SIE	17	existir	0	5	1	R
#,E,X,I,S,T,I,M,O,S,$	2	existimos	1PIE	17	existir	0	5	4	R
#,E,X,I,S,T,í,S,$	2	existís	2PIE	17	existir	0	5	2	R
#,E,X,I,S,T,E,N,$	45	existen	3PIE	17	existir	0	5	2	R
#,E,X,I,S,T,í,$	2	existí	1SIA	17	existir	0	5	1	R
#,E,X,I,S,T,I,S,T,E,$	2	exististe	2SIA	17	existir	0	5	4	R
#,E,X,I,S,T,I,ó,$	3	existió	3SIA	17	existir	0	5	2	R
#,E,X,I,S,T,I,M,O,S,$	1	existimos	1PIA	17	existir	0	5	4	R
#,E,X,I,S,T,I,S,T,E,I,S,$	1	exististeis	2PIA	17	existir	0	5	6	R
#,E,X,I,S,T,I,E,R,O,N,$	2	existieron	3PIA	17	existir	0	5	5	R
#,G,U,S,T,A,R,$	4	gustar	i	18	gustar	0	4	2	R
#,G,U,S,T,A,N,D,O,$	2	gustando	pE	18	gustar	0	4	4	R
#,G,U,S,T,A,D,O,$	14	gustado	pA	18	gustar	0	4	3	R
#,G,U,S,T,O,$	3	gusto	1SIE	18	gustar	0	4	1	R
#,G,U,S,T,A,S,$	2	gustas	2SIE	18	gustar	0	4	2	R
#,G,U,S,T,A,$	62	gusta	3SIE	18	gustar	0	4	1	R
#,G,U,S,T,A,M,O,S,$	1	gustamos	1PIE	18	gustar	0	4	4	R
#,G,U,S,T,á,I,S,$	2	gustáis	2PIE	18	gustar	0	4	3	R
#,G,U,S,T,A,N,$	16	gustan	3PIE	18	gustar	0	4	2	R
#,G,U,S,T,E,',$	2	gusté	1SIA	18	gustar	0	4	2	R
#,G,U,S,T,A,S,T,E,$	2	gustaste	2SIA	18	gustar	0	4	4	R
#,G,U,S,T,ó,$	8	gustó	3SIA	18	gustar	0	4	1	R
#,G,U,S,T,A,M,O,S,$	2	gustamos	1PIA	18	gustar	0	4	4	R
#,G,U,S,T,A,S,T,E,I,S,$	2	gustasteis	2PIA	18	gustar	0	4	6	R
#,G,U,S,T,A,R,O,N,$	3	gustaron	3PIA	18	gustar	0	4	4	R
#,H,A,B,E,R,$	59	haber	i	19	haber	0	3	2	I
#,H,A,B,I,E,N,D,O,$	6	habiendo	pE	19	haber	0	3	5	I
#,H,A,B,I,D,O,$	10	habido	pA	19	haber	0	3	3	I
#,H,E,$	116	he	1SIE	19	haber	0	1	1	I
#,H,A,S,$	23	has	2SIE	19	haber	0	1	2	I
#,H,A,$	1001	ha	3SIE	19	haber	0	1	1	I
#,H,E,M,O,S,$	72	hemos	1PIE	19	haber	0	1	4	I
#,H,A,B,E,',I,S,$	6	habéis	2PIE	19	haber	0	3	4	I
#,H,A,N,$	305	han	3PIE	19	haber	0	1	2	I
#,H,U,B,E,$	2	hube	1SIA	19	haber	0	3	1	I
#,H,U,B,I,S,T,E,$	2	hubiste	2SIA	19	haber	0	3	4	I

#,H,U,B,O,$	15	hubo	3SIA	19	haber	0	3	1	I
#,H,U,B,I,M,O,S,$	2	hubimos	1PIA	19	haber	0	3	4	I
#,H,U,B,I,S,T,E,I,S,$	2	hubisteis	2PIA	19	haber	0	3	6	I
#,H,U,B,I,E,R,O,N,$	2	hubieron	3PIA	19	haber	0	3	5	I
#,H,A,B,L,A,R,$	55	hablar	i	20	hablar	0	4	2	R
#,H,A,B,L,A,N,D,O,$	26	hablando	pE	20	hablar	0	4	4	R
#,H,A,B,L,A,D,O,$	13	hablado	pA	20	hablar	0	4	3	R
#,H,A,B,L,O,$	9	hablo	1SIE	20	hablar	0	4	1	R
#,H,A,B,L,A,S,$	2	hablas	2SIE	20	hablar	0	4	2	R
#,H,A,B,L,A,$	31	habla	3SIE	20	hablar	0	4	1	R
#,H,A,B,L,A,M,O,S,$	14	hablamos	1PIE	20	hablar	0	4	4	R
#,H,A,B,L,á,I,S,$	2	habláis	2PIE	20	hablar	0	4	3	R
#,H,A,B,L,A,N,$	12	hablan	3PIE	20	hablar	0	4	2	R
#,H,A,B,L,E,',$	3	hablé	1SIA	20	hablar	0	4	2	R
#,H,A,B,L,A,S,T,E,$	2	hablaste	2SIA	20	hablar	0	4	4	R
#,H,A,B,L,ó,$	7	habló	3SIA	20	hablar	0	4	1	R
#,H,A,B,L,A,M,O,S,$	1	hablamos	1PIA	20	hablar	0	4	4	R
#,H,A,B,L,A,S,T,E,I,S,$	2	hablasteis	2PIA	20	hablar	0	4	6	R
#,H,A,B,L,A,R,O,N,$	3	hablaron	3PIA	20	hablar	0	4	4	R
#,H,A,C,E,R,$	286	hacer	i	21	hacer	0	3	2	I
#,H,A,C,I,E,N,D,O,$	62	haciendo	pE	21	hacer	0	3	5	I
#,H,E,C,H,O,$	115	hecho	pA	21	hacer	0	2	3	I
#,H,A,G,O,$	20	hago	1SIE	21	hacer	0	3	1	I
#,H,A,C,E,S,$	5	haces	2SIE	21	hacer	0	3	2	I
#,H,A,C,E,$	328	hace	3SIE	21	hacer	0	3	1	I
#,H,A,C,E,M,O,S,$	18	hacemos	1PIE	21	hacer	0	3	4	I
#,H,A,C,E,',I,S,$	3	hacéis	2PIE	21	hacer	0	3	4	I
#,H,A,C,E,N,$	74	hacen	3PIE	21	hacer	0	3	2	I
#,H,I,C,E,$	13	hice	1SIA	21	hacer	0	3	1	I
#,H,I,C,I,S,T,E,$	3	hiciste	2SIA	21	hacer	0	3	4	I
#,H,I,Z,O,$	64	hizo	3SIA	21	hacer	0	3	1	I
#,H,I,C,I,M,O,S,$	6	hicimos	1PIA	21	hacer	0	3	4	I
#,H,I,C,I,S,T,E,I,S,$	2	hicisteis	2PIA	21	hacer	0	3	6	I
#,H,I,C,I,E,R,O,N,$	23	hicieron	3PIA	21	hacer	0	3	5	I
#,I,N,C,L,U,I,R,$	17	incluir	i	22	incluir	0	5	2	I
#,I,N,C,L,U,Y,E,N,D,O,$	21	incluyendo	pE	22	incluir	0	5	5	I
#,I,N,C,L,U,I,D,O,$	20	incluido	pA	22	incluir	0	5	3	I
#,I,N,C,L,U,Y,O,$	3	incluyo	1SIE	22	incluir	0	6	1	I
#,I,N,C,L,U,Y,E,S,$	2	incluyes	2SIE	22	incluir	0	6	2	I
#,I,N,C,L,U,Y,E,$	44	incluye	3SIE	22	incluir	0	6	1	I
#,I,N,C,L,U,I,M,O,S,$	1	incluimos	1PIE	22	incluir	0	5	4	I
#,I,N,C,L,U,í,S,$	2	incluís	2PIE	22	incluir	0	5	2	I
#,I,N,C,L,U,Y,E,N,$	20	incluyen	3PIE	22	incluir	0	6	2	I
#,I,N,C,L,U,í,$	2	incluí	1SIA	22	incluir	0	5	1	I
#,I,N,C,L,U,I,S,T,E,$	2	incluiste	2SIA	22	incluir	0	5	4	I
#,I,N,C,L,U,Y,ó,$	4	incluyó	3SIA	22	incluir	0	6	1	I
#,I,N,C,L,U,I,M,O,S,$	2	incluimos	1PIA	22	incluir	0	5	4	I
#,I,N,C,L,U,I,S,T,E,I,S,$	2	incluisteis	2PIA	22	incluir	0	5	6	I

#,I,N,C,L,U,Y,E,R,O,N,$	2	incluyeron	3PIA	22	incluir	0	6	4	I
#,I,R,$	91	ir	i	23	ir	0	1	1	I
#,Y,E,N,D,O,$	4	yendo	pE	23	ir	0	1	4	I
#,I,D,O,$	32	ido	pA	23	ir	0	1	2	I
#,V,O,Y,$	59	voy	1SIE	23	ir	0	1	2	I
#,V,A,S,$	19	vas	2SIE	23	ir	0	1	2	I
#,V,A,$	185	va	3SIE	23	ir	0	1	1	I
#,V,A,M,O,S,$	63	vamos	1PIE	23	ir	0	1	4	I
#,V,á,I,S,$	6	váis	2PIE	23	ir	0	1	3	I
#,V,A,N,$	91	van	3PIE	23	ir	0	1	2	I
#,F,U,I,$	7	fui	1SIA	23	ir	0	2	1	I
#,F,U,I,S,T,E,$	2	fuiste	2SIA	23	ir	0	2	4	I
#,F,U,E,$	12	fue	3SIA	23	ir	0	2	1	I
#,F,U,I,M,O,S,$	5	fuimos	1PIA	23	ir	0	2	4	I
#,F,U,I,S,T,E,I,S,$	2	fuisteis	2PIA	23	ir	0	2	6	I
#,F,U,E,R,O,N,$	9	fueron	3PIA	23	ir	0	2	4	I
#,L,L,A,M,A,R,$	17	llamar	i	24	llamar	0	4	2	R
#,L,L,A,M,A,N,D,O,$	5	llamando	pE	24	llamar	0	4	4	R
#,L,L,A,M,A,D,O,$	31	llamado	pA	24	llamar	0	4	3	R
#,L,L,A,M,O,$	7	llamo	1SIE	24	llamar	0	4	1	R
#,L,L,A,M,A,S,$	2	llamas	2SIE	24	llamar	0	4	2	R
#,L,L,A,M,A,$	30	llama	3SIE	24	llamar	0	4	1	R
#,L,L,A,M,A,M,O,S,$	5	llamamos	1PIE	24	llamar	0	4	4	R
#,L,L,A,M,á,I,S,$	2	llamáis	2PIE	24	llamar	0	4	3	R
#,L,L,A,M,A,N,$	11	llaman	3PIE	24	llamar	0	4	2	R
#,L,L,A,M,E,',$	2	llamé	1SIA	24	llamar	0	4	2	R
#,L,L,A,M,A,S,T,E,$	2	llamaste	2SIA	24	llamar	0	4	4	R
#,L,L,A,M,ó,$	8	llamó	3SIA	24	llamar	0	4	1	R
#,L,L,A,M,A,M,O,S,$	1	llamamos	1PIA	24	llamar	0	4	4	R
#,L,L,A,M,A,S,T,E,I,S,$	2	llamasteis	2PIA	24	llamar	0	4	6	R
#,L,L,A,M,A,R,O,N,$	3	llamaron	3PIA	24	llamar	0	4	4	R
#,L,L,E,G,A,R,$	81	llegar	i	25	llegar	0	4	2	R
#,L,L,E,G,A,N,D,O,$	11	llegando	pE	25	llegar	0	4	4	R
#,L,L,E,G,A,D,O,$	34	llegado	pA	25	llegar	0	4	3	R
#,L,L,E,G,O,$	7	llego	1SIE	25	llegar	0	4	1	R
#,L,L,E,G,A,S,$	3	llegas	2SIE	25	llegar	0	4	2	R
#,L,L,E,G,A,$	45	llega	3SIE	25	llegar	0	4	1	R
#,L,L,E,G,A,M,O,S,$	8	llegamos	1PIE	25	llegar	0	4	4	R
#,L,L,E,G,á,I,S,$	2	llegáis	2PIE	25	llegar	0	4	3	R
#,L,L,E,G,A,N,$	18	llegan	3PIE	25	llegar	0	4	2	R
#,L,L,E,G,U,E,',$	5	llegué	1SIA	25	llegar	0	4	3	R
#,L,L,E,G,A,S,T,E,$	2	llegaste	2SIA	25	llegar	0	4	4	R
#,L,L,E,G,ó,$	32	llegó	3SIA	25	llegar	0	4	1	R
#,L,L,E,G,A,M,O,S,$	1	llegamos	1PIA	25	llegar	0	4	4	R
#,L,L,E,G,A,S,T,E,I,S,$	2	llegasteis	2PIA	25	llegar	0	4	6	R
#,L,L,E,G,A,R,O,N,$	12	llegaron	3PIA	25	llegar	0	4	4	R
#,L,L,E,V,A,R,$	57	llevar	i	26	llevar	0	4	2	R
#,L,L,E,V,A,N,D,O,$	11	llevando	pE	26	llevar	0	4	4	R

#,L,L,E,V,A,D,O,$	25	llevado	pA	26	llevar	0	4	3	R
#,L,L,E,V,O,$	18	llevo	1SIE	26	llevar	0	4	1	R
#,L,L,E,V,A,S,$	6	llevas	2SIE	26	llevar	0	4	2	R
#,L,L,E,V,A,$	8	lleva	3SIE	26	llevar	0	4	1	R
#,L,L,E,V,A,M,O,S,$	10	llevamos	1PIE	26	llevar	0	4	4	R
#,L,L,E,V,á,I,S,$	2	lleváis	2PIE	26	llevar	0	4	3	R
#,L,L,E,V,A,N,$	26	llevan	3PIE	26	llevar	0	4	2	R
#,L,L,E,V,E,',$	3	llevé	1SIA	26	llevar	0	4	2	R
#,L,L,E,V,A,S,T,E,$	2	llevaste	2SIA	26	llevar	0	4	4	R
#,L,L,E,V,ó,$	17	llevó	3SIA	26	llevar	0	4	1	R
#,L,L,E,V,A,M,O,S,$	1	llevamos	1PIA	26	llevar	0	4	4	R
#,L,L,E,V,A,S,T,E,I,S,$	2	llevasteis	2PIA	26	llevar	0	4	6	R
#,L,L,E,V,A,R,O,N,$	8	llevaron	3PIA	26	llevar	0	4	4	R
#,O,F,R,E,C,E,R,$	33	ofrecer	i	27	ofrecer	0	5	2	I
#,O,F,R,E,C,I,E,N,D,O,$	10	ofreciendo	pE	27	ofrecer	0	5	5	I
#,O,F,R,E,C,I,D,O,$	7	ofrecido	pA	27	ofrecer	0	5	3	I
#,O,F,R,E,Z,C,O,$	3	ofrezco	1SIE	27	ofrecer	0	6	1	I
#,O,F,R,E,C,E,S,$	2	ofreces	2SIE	27	ofrecer	0	5	2	I
#,O,F,R,E,C,E,$	81	ofrece	3SIE	27	ofrecer	0	5	1	I
#,O,F,R,E,C,E,M,O,S,$	9	ofrecemos	1PIE	27	ofrecer	0	5	4	I
#,O,F,R,E,C,E,',I,S,$	2	ofrecéis	2PIE	27	ofrecer	0	5	4	I
#,O,F,R,E,C,E,N,$	25	ofrecen	3PIE	27	ofrecer	0	5	2	I
#,O,F,R,E,C,í,$	2	ofrecí	1SIA	27	ofrecer	0	5	1	I
#,O,F,R,E,C,I,S,T,E,$	2	ofreciste	2SIA	27	ofrecer	0	5	4	I
#,O,F,R,E,C,I,ó,$	7	ofreció	3SIA	27	ofrecer	0	5	2	I
#,O,F,R,E,C,I,M,O,S,$	2	ofrecimos	1PIA	27	ofrecer	0	5	4	I
#,O,F,R,E,C,I,S,T,E,I,S,$	2	ofrecisteis	2PIA	27	ofrecer	0	5	6	I
#,O,F,R,E,C,I,E,R,O,N,$	4	ofrecieron	3PIA	27	ofrecer	0	5	5	I
#,P,A,R,E,C,E,R,$	18	parecer	i	28	parecer	0	5	2	I
#,P,A,R,E,C,I,E,N,D,O,$	2	pareciendo	pE	28	parecer	0	5	5	I
#,P,A,R,E,C,I,D,O,$	18	parecido	pA	28	parecer	0	5	3	I
#,P,A,R,E,Z,C,O,$	2	parezco	1SIE	28	parecer	0	6	1	I
#,P,A,R,E,C,E,S,$	2	pareces	2SIE	28	parecer	0	5	2	I
#,P,A,R,E,C,E,$	159	parece	3SIE	28	parecer	0	5	1	I
#,P,A,R,E,C,E,M,O,S,$	2	parecemos	1PIE	28	parecer	0	5	4	I
#,P,A,R,E,C,E,',I,S,$	2	parecéis	2PIE	28	parecer	0	5	4	I
#,P,A,R,E,C,E,N,$	20	parecen	3PIE	28	parecer	0	5	2	I
#,P,A,R,E,C,í,$	2	parecí	1SIA	28	parecer	0	5	1	I
#,P,A,R,E,C,I,S,T,E,$	2	pareciste	2SIA	28	parecer	0	5	4	I
#,P,A,R,E,C,I,ó,$	8	pareció	3SIA	28	parecer	0	5	2	I
#,P,A,R,E,C,I,M,O,S,$	2	parecimos	1PIA	28	parecer	0	5	4	I
#,P,A,R,E,C,I,S,T,E,I,S,$	2	parecisteis	2PIA	28	parecer	0	5	6	I
#,P,A,R,E,C,I,E,R,O,N,$	2	parecieron	3PIA	28	parecer	0	5	5	I
#,P,A,S,A,R,$	59	pasar	i	29	pasar	0	3	2	R
#,P,A,S,A,N,D,O,$	24	pasando	pE	29	pasar	0	3	4	R
#,P,A,S,A,D,O,$	128	pasado	pA	29	pasar	0	3	3	R
#,P,A,S,O,$	9	paso	1SIE	29	pasar	0	3	1	R
#,P,A,S,A,S,$	2	pasas	2SIE	29	pasar	0	3	2	R

#,P,A,S,A,$	63	pasa	3SIE	29	pasar	0	3	1	R
#,P,A,S,A,M,O,S,$	8	pasamos	1PIE	29	pasar	0	3	4	R
#,P,A,S,á,I,S,$	2	pasáis	2PIE	29	pasar	0	3	3	R
#,P,A,S,A,N,$	17	pasan	3PIE	29	pasar	0	3	2	R
#,P,A,S,E,',$	4	pasé	1SIA	29	pasar	0	3	2	R
#,P,A,S,A,S,T,E,$	2	pasaste	2SIA	29	pasar	0	3	4	R
#,P,A,S,ó,$	21	pasó	3SIA	29	pasar	0	3	1	R
#,P,A,S,A,M,O,S,$	1	pasamos	1PIA	29	pasar	0	3	4	R
#,P,A,S,A,S,T,E,I,S,$	2	pasasteis	2PIA	29	pasar	0	3	6	R
#,P,A,S,A,R,O,N,$	7	pasaron	3PIA	29	pasar	0	3	4	R
#,P,E,N,S,A,R,$	42	pensar	i	30	pensar	0	4	2	I
#,P,E,N,S,A,N,D,O,$	18	pensando	pE	30	pensar	0	4	4	I
#,P,E,N,S,A,D,O,$	14	pensado	pA	30	pensar	0	4	3	I
#,P,I,E,N,S,O,$	21	pienso	1SIE	30	pensar	0	5	1	I
#,P,I,E,N,S,A,S,$	5	piensas	2SIE	30	pensar	0	5	2	I
#,P,I,E,N,S,A,$	17	piensa	3SIE	30	pensar	0	5	1	I
#,P,E,N,S,A,M,O,S,$	8	pensamos	1PIE	30	pensar	0	4	4	I
#,P,E,N,S,á,I,S,$	2	pensáis	2PIE	30	pensar	0	4	3	I
#,P,I,E,N,S,A,N,$	9	piensan	3PIE	30	pensar	0	5	2	I
#,P,E,N,S,E,',$	7	pensé	1SIA	30	pensar	0	4	2	I
#,P,E,N,S,A,S,T,E,$	2	pensaste	2SIA	30	pensar	0	4	4	I
#,P,E,N,S,ó,$	4	pensó	3SIA	30	pensar	0	4	1	I
#,P,E,N,S,A,M,O,S,$	1	pensamos	1PIA	30	pensar	0	4	4	I
#,P,E,N,S,A,S,T,E,I,S,$	2	pensasteis	2PIA	30	pensar	0	4	6	I
#,P,E,N,S,A,R,O,N,$	2	pensaron	3PIA	30	pensar	0	4	4	I
#,P,E,R,M,I,T,I,R,$	18	permitir	i	31	permitir	0	6	2	R
#,P,E,R,M,I,T,I,E,N,D,O,$	8	permitiendo	pE	31	permitir	0	6	5	R
#,P,E,R,M,I,T,I,D,O,$	15	permitido	pA	31	permitir	0	6	3	R
#,P,E,R,M,I,T,O,$	2	permito	1SIE	31	permitir	0	6	1	R
#,P,E,R,M,I,T,E,S,$	2	permites	2SIE	31	permitir	0	6	2	R
#,P,E,R,M,I,T,E,$	82	permite	3SIE	31	permitir	0	6	1	R
#,P,E,R,M,I,T,I,M,O,S,$	2	permitimos	1PIE	31	permitir	0	6	4	R
#,P,E,R,M,I,T,í,S,$	2	permitís	2PIE	31	permitir	0	6	2	R
#,P,E,R,M,I,T,E,N,$	25	permiten	3PIE	31	permitir	0	6	2	R
#,P,E,R,M,I,T,í,$	2	permití	1SIA	31	permitir	0	6	1	R
#,P,E,R,M,I,T,I,S,T,E,$	2	permitiste	2SIA	31	permitir	0	6	4	R
#,P,E,R,M,I,T,I,ó,$	8	permitió	3SIA	31	permitir	0	6	2	R
#,P,E,R,M,I,T,I,M,O,S,$	1	permitimos	1PIA	31	permitir	0	6	4	R
#,P,E,R,M,I,T,I,S,T,E,I,S,$	2	permitisteis	2PIA	31	permitir	0	6	6	R
#,P,E,R,M,I,T,I,E,R,O,N,$	4	permitieron	3PIA	31	permitir	0	6	5	R
#,P,O,D,E,R,$	91	poder	i	32	poder	0	3	2	I
#,P,U,D,I,E,N,D,O,$	11	pudiendo	pE	32	poder	0	3	5	I
#,P,O,D,I,D,O,$	33	podido	pA	32	poder	0	3	3	I
#,P,U,E,D,O,$	54	puedo	1SIE	32	poder	0	4	1	I
#,P,U,E,D,E,S,$	68	puedes	2SIE	32	poder	0	4	2	I
#,P,U,E,D,E,$	421	puede	3SIE	32	poder	0	4	1	I
#,P,O,D,E,M,O,S,$	85	podemos	1PIE	32	poder	0	3	4	I
#,P,O,D,E,',I,S,$	11	podéis	2PIE	32	poder	0	3	4	I

#,P,U,E,D,E,N,$	194	pueden	3PIE	32	poder	0	4	2	I
#,P,U,D,E,$	9	pude	1SIA	32	poder	0	3	1	I
#,P,U,D,I,S,T,E,$	2	pudiste	2SIA	32	poder	0	3	4	I
#,P,U,D,O,$	24	pudo	3SIA	32	poder	0	3	1	I
#,P,U,D,I,M,O,S,$	6	pudimos	1PIA	32	poder	0	3	4	I
#,P,U,D,I,S,T,E,I,S,$	2	pudisteis	2PIA	32	poder	0	3	6	I
#,P,U,D,I,E,R,O,N,$	10	pudieron	3PIA	32	poder	0	3	5	I
#,P,O,N,E,R,$	71	poner	i	33	poner	0	3	2	I
#,P,O,N,I,E,N,D,O,$	13	poniendo	pE	33	poner	0	3	5	I
#,P,U,E,S,T,O,$	57	puesto	pA	33	poner	0	4	2	I
#,P,O,N,G,O,$	13	pongo	1SIE	33	poner	0	4	1	I
#,P,O,N,E,S,$	6	pones	2SIE	33	poner	0	3	2	I
#,P,O,N,E,$	44	pone	3SIE	33	poner	0	3	1	I
#,P,O,N,E,M,O,S,$	9	ponemos	1PIE	33	poner	0	3	4	I
#,P,O,N,E,',I,S,$	2	ponéis	2PIE	33	poner	0	3	4	I
#,P,O,N,E,N,$	20	ponen	3PIE	33	poner	0	3	2	I
#,P,U,S,E,$	7	puse	1SIA	33	poner	0	3	1	I
#,P,U,S,I,S,T,E,$	2	pusiste	2SIA	33	poner	0	3	4	I
#,P,U,S,O,$	21	puso	3SIA	33	poner	0	3	1	I
#,P,U,S,I,M,O,S,$	3	pusimos	1PIA	33	poner	0	3	4	I
#,P,U,S,I,S,T,E,I,S,$	2	pusisteis	2PIA	33	poner	0	3	6	I
#,P,U,S,I,E,R,O,N,$	8	pusieron	3PIA	33	poner	0	3	5	I
#,P,R,E,S,E,N,T,A,R,$	34	presentar	i	34	presentar	0	7	2	R
#,P,R,E,S,E,N,T,A,N,D,O,$	6	presentando	pE	34	presentar	0	7	4	R
#,P,R,E,S,E,N,T,A,D,O,$	32	presentado	pA	34	presentar	0	7	3	R
#,P,R,E,S,E,N,T,O,$	4	presento	1SIE	34	presentar	0	7	1	R
#,P,R,E,S,E,N,T,A,S,$	2	presentas	2SIE	34	presentar	0	7	2	R
#,P,R,E,S,E,N,T,A,$	48	presenta	3SIE	34	presentar	0	7	1	R
#,P,R,E,S,E,N,T,A,M,O,S,$	7	presentamos	1PIE	34	presentar	0	7	4	R
#,P,R,E,S,E,N,T,á,I,S,$	2	presentáis	2PIE	34	presentar	0	7	3	R
#,P,R,E,S,E,N,T,A,N,$	21	presentan	3PIE	34	presentar	0	7	2	R
#,P,R,E,S,E,N,T,E,',$	2	presenté	1SIA	34	presentar	0	7	2	R
#,P,R,E,S,E,N,T,A,S,T,E,$	2	presentaste	2SIA	34	presentar	0	7	4	R
#,P,R,E,S,E,N,T,ó,$	17	presentó	3SIA	34	presentar	0	7	1	R
#,P,R,E,S,E,N,T,A,M,O,S,$	1	presentamos	1PIA	34	presentar	0	7	4	R
#,P,R,E,S,E,N,T,A,S,T,E,I,S,$	2	presentasteis	2PIA	34	presentar	0	7	6	R
#,P,R,E,S,E,N,T,A,R,O,N,$	7	presentaron	3PIA	34	presentar	0	7	4	R
#,P,U,B,L,I,C,A,R,$	11	publicar	i	35	publicar	0	6	2	R
#,P,U,B,L,I,C,A,N,D,O,$	3	publicando	pE	35	publicar	0	6	4	R
#,P,U,B,L,I,C,A,D,O,$	62	publicado	pA	35	publicar	0	6	3	R
#,P,U,B,L,I,C,O,$	7	publico	1SIE	35	publicar	0	6	1	R
#,P,U,B,L,I,C,A,S,$	3	publicas	2SIE	35	publicar	0	6	2	R
#,P,U,B,L,I,C,A,$	15	publica	3SIE	35	publicar	0	6	1	R
#,P,U,B,L,I,C,A,M,O,S,$	2	publicamos	1PIE	35	publicar	0	6	4	R
#,P,U,B,L,I,C,á,I,S,$	2	publicáis	2PIE	35	publicar	0	6	3	R
#,P,U,B,L,I,C,A,N,$	5	publican	3PIE	35	publicar	0	6	2	R
#,P,U,B,L,I,Q,U,E,',$	2	publiqué	1SIA	35	publicar	0	6	3	R
#,P,U,B,L,I,C,A,S,T,E,$	2	publicaste	2SIA	35	publicar	0	6	4	R

#,P,U,B,L,I,C,ó,$	9	publicó	3SIA	35	publicar	0	6	1	R
#,P,U,B,L,I,C,A,M,O,S,$	1	publicamos	1PIA	35	publicar	0	6	4	R
#,P,U,B,L,I,C,A,S,T,E,I,S,$	2	publicasteis	2PIA	35	publicar	0	6	6	R
#,P,U,B,L,I,C,A,R,O,N,$	3	publicaron	3PIA	35	publicar	0	6	4	R
#,Q,U,E,D,A,R,$	17	quedar	i	36	quedar	0	4	2	R
#,Q,U,E,D,A,N,D,O,$	9	quedando	pE	36	quedar	0	4	4	R
#,Q,U,E,D,A,D,O,$	24	quedado	pA	36	quedar	0	4	3	R
#,Q,U,E,D,O,$	10	quedo	1SIE	36	quedar	0	4	1	R
#,Q,U,E,D,A,S,$	2	quedas	2SIE	36	quedar	0	4	2	R
#,Q,U,E,D,A,$	67	queda	3SIE	36	quedar	0	4	1	R
#,Q,U,E,D,A,M,O,S,$	6	quedamos	1PIE	36	quedar	0	4	4	R
#,Q,U,E,D,á,I,S,$	2	quedáis	2PIE	36	quedar	0	4	3	R
#,Q,U,E,D,A,N,$	29	quedan	3PIE	36	quedar	0	4	2	R
#,Q,U,E,D,E,',$	5	quedé	1SIA	36	quedar	0	4	2	R
#,Q,U,E,D,A,S,T,E,$	2	quedaste	2SIA	36	quedar	0	4	4	R
#,Q,U,E,D,ó,$	19	quedó	3SIA	36	quedar	0	4	1	R
#,Q,U,E,D,A,M,O,S,$	1	quedamos	1PIA	36	quedar	0	4	4	R
#,Q,U,E,D,A,S,T,E,I,S,$	2	quedasteis	2PIA	36	quedar	0	4	6	R
#,Q,U,E,D,A,R,O,N,$	8	quedaron	3PIA	36	quedar	0	4	4	R
#,Q,U,E,R,E,R,$	11	querer	i	37	querer	0	4	2	I
#,Q,U,E,R,I,E,N,D,O,$	3	queriendo	pE	37	querer	0	4	5	I
#,Q,U,E,R,I,D,O,$	21	querido	pA	37	querer	0	4	3	I
#,Q,U,I,E,R,O,$	56	quiero	1SIE	37	querer	0	5	1	I
#,Q,U,I,E,R,E,S,$	29	quieres	2SIE	37	querer	0	5	2	I
#,Q,U,I,E,R,E,$	72	quiere	3SIE	37	querer	0	5	1	I
#,Q,U,E,R,E,M,O,S,$	34	queremos	1PIE	37	querer	0	4	4	I
#,Q,U,E,R,E,',I,S,$	5	queréis	2PIE	37	querer	0	4	4	I
#,Q,U,I,E,R,E,N,$	33	quieren	3PIE	37	querer	0	5	2	I
#,Q,U,I,S,E,$	4	quise	1SIA	37	querer	0	4	1	I
#,Q,U,I,S,I,S,T,E,$	2	quisiste	2SIA	37	querer	0	4	4	I
#,Q,U,I,S,O,$	11	quiso	3SIA	37	querer	0	4	1	I
#,Q,U,I,S,I,M,O,S,$	2	quisimos	1PIA	37	querer	0	4	4	I
#,Q,U,I,S,I,S,T,E,I,S,$	2	quisisteis	2PIA	37	querer	0	4	6	I
#,Q,U,I,S,I,E,R,O,N,$	4	quisieron	3PIA	37	querer	0	4	5	I
#,R,E,A,L,I,Z,A,R,$	92	realizar	i	38	realizar	0	6	2	R
#,R,E,A,L,I,Z,A,N,D,O,$	17	realizando	pE	38	realizar	0	6	4	R
#,R,E,A,L,I,Z,A,D,O,$	48	realizado	pA	38	realizar	0	6	3	R
#,R,E,A,L,I,Z,O,$	3	realizo	1SIE	38	realizar	0	6	1	R
#,R,E,A,L,I,Z,A,S,$	2	realizas	2SIE	38	realizar	0	6	2	R
#,R,E,A,L,I,Z,A,$	33	realiza	3SIE	38	realizar	0	6	1	R
#,R,E,A,L,I,Z,A,M,O,S,$	5	realizamos	1PIE	38	realizar	0	6	4	R
#,R,E,A,L,I,Z,á,I,S,$	2	realizáis	2PIE	38	realizar	0	6	3	R
#,R,E,A,L,I,Z,A,N,$	19	realizan	3PIE	38	realizar	0	6	2	R
#,R,E,A,L,I,C,E,',$	2	realicé	1SIA	38	realizar	0	6	2	R
#,R,E,A,L,I,Z,A,S,T,E,$	2	realizaste	2SIA	38	realizar	0	6	4	R
#,R,E,A,L,I,Z,ó,$	16	realizó	3SIA	38	realizar	0	6	1	R
#,R,E,A,L,I,Z,A,M,O,S,$	1	realizamos	1PIA	38	realizar	0	6	4	R
#,R,E,A,L,I,Z,A,S,T,E,I,S,$	2	realizasteis	2PIA	38	realizar	0	6	6	R

#,R,E,A,L,I,Z,A,R,O,N,$	7	realizaron	3PIA	38	realizar	0	6	4	R
#,R,E,C,I,B,I,R,$	28	recibir	i	39	recibir	0	5	2	R
#,R,E,C,I,B,I,E,N,D,O,$	6	recibiendo	pE	39	recibir	0	5	5	R
#,R,E,C,I,B,I,D,O,$	28	recibido	pA	39	recibir	0	5	3	R
#,R,E,C,I,B,O,$	2	recibo	1SIE	39	recibir	0	5	1	R
#,R,E,C,I,B,E,S,$	2	recibes	2SIE	39	recibir	0	5	2	R
#,R,E,C,I,B,E,$	18	recibe	3SIE	39	recibir	0	5	1	R
#,R,E,C,I,B,I,M,O,S,$	4	recibimos	1PIE	39	recibir	0	5	4	R
#,R,E,C,I,B,í,S,$	2	recibís	2PIE	39	recibir	0	5	2	R
#,R,E,C,I,B,E,N,$	11	reciben	3PIE	39	recibir	0	5	2	R
#,R,E,C,I,B,í,$	3	recibí	1SIA	39	recibir	0	5	1	R
#,R,E,C,I,B,I,S,T,E,$	2	recibiste	2SIA	39	recibir	0	5	4	R
#,R,E,C,I,B,I,ó,$	15	recibió	3SIA	39	recibir	0	5	2	R
#,R,E,C,I,B,I,M,O,S,$	1	recibimos	1PIA	39	recibir	0	5	4	R
#,R,E,C,I,B,I,S,T,E,I,S,$	2	recibisteis	2PIA	39	recibir	0	5	6	R
#,R,E,C,I,B,I,E,R,O,N,$	6	recibieron	3PIA	39	recibir	0	5	5	R
#,S,A,B,E,R,$	76	saber	i	40	saber	0	3	2	I
#,S,A,B,I,E,N,D,O,$	7	sabiendo	pE	40	saber	0	3	5	I
#,S,A,B,I,D,O,$	12	sabido	pA	40	saber	0	3	3	I
#,S,E,',$	56	sé	1SIE	40	saber	0	1	2	I
#,S,A,B,E,S,$	21	sabes	2SIE	40	saber	0	3	2	I
#,S,A,B,E,$	59	sabe	3SIE	40	saber	0	3	1	I
#,S,A,B,E,M,O,S,$	30	sabemos	1PIE	40	saber	0	3	4	I
#,S,A,B,E,',I,S,$	7	sabéis	2PIE	40	saber	0	3	4	I
#,S,A,B,E,N,$	24	saben	3PIE	40	saber	0	3	2	I
#,S,U,P,E,$	3	supe	1SIA	40	saber	0	3	1	I
#,S,U,P,I,S,T,E,$	2	supiste	2SIA	40	saber	0	3	4	I
#,S,U,P,O,$	7	supo	3SIA	40	saber	0	3	1	I
#,S,U,P,I,M,O,S,$	2	supimos	1PIA	40	saber	0	3	4	I
#,S,U,P,I,S,T,E,I,S,$	2	supisteis	2PIA	40	saber	0	3	6	I
#,S,U,P,I,E,R,O,N,$	3	supieron	3PIA	40	saber	0	3	5	I
#,S,A,L,I,R,$	45	salir	i	41	salir	0	3	2	I
#,S,A,L,I,E,N,D,O,$	7	saliendo	pE	41	salir	0	3	5	I
#,S,A,L,I,D,O,$	16	salido	pA	41	salir	0	3	3	I
#,S,A,L,G,O,$	4	salgo	1SIE	41	salir	0	4	1	I
#,S,A,L,E,S,$	2	sales	2SIE	41	salir	0	3	2	I
#,S,A,L,E,$	32	sale	3SIE	41	salir	0	3	1	I
#,S,A,L,I,M,O,S,$	5	salimos	1PIE	41	salir	0	3	4	I
#,S,A,L,í,S,$	2	salís	2PIE	41	salir	0	3	2	I
#,S,A,L,E,N,$	15	salen	3PIE	41	salir	0	3	2	I
#,S,A,L,í,$	3	salí	1SIA	41	salir	0	3	1	I
#,S,A,L,I,S,T,E,$	2	saliste	2SIA	41	salir	0	3	4	I
#,S,A,L,I,ó,$	15	salió	3SIA	41	salir	0	3	2	I
#,S,A,L,I,M,O,S,$	1	salimos	1PIA	41	salir	0	3	4	I
#,S,A,L,I,S,T,E,I,S,$	2	salisteis	2PIA	41	salir	0	3	6	I
#,S,A,L,I,E,R,O,N,$	6	salieron	3PIA	41	salir	0	3	5	I
#,S,E,G,U,I,R,$	80	seguir	i	42	seguir	0	4	2	I
#,S,I,G,U,I,E,N,D,O,$	19	siguiendo	pE	42	seguir	0	4	5	I

#,S,E,G,U,I,D,O,$	17	seguido	pA	42	seguir	0	4	3	I
#,S,I,G,O,$	14	sigo	1SIE	42	seguir	0	3	1	I
#,S,I,G,U,E,S,$	4	sigues	2SIE	42	seguir	0	4	2	I
#,S,I,G,U,E,$	73	sigue	3SIE	42	seguir	0	4	1	I
#,S,E,G,U,I,M,O,S,$	2	seguimos	1PIE	42	seguir	0	4	4	I
#,S,E,G,U,í,S,$	2	seguís	2PIE	42	seguir	0	4	2	I
#,S,I,G,U,E,N,$	29	siguen	3PIE	42	seguir	0	4	2	I
#,S,E,G,U,í,$	2	seguí	1SIA	42	seguir	0	4	1	I
#,S,E,G,U,I,S,T,E,$	2	seguiste	2SIA	42	seguir	0	4	4	I
#,S,I,G,U,I,ó,$	7	siguió	3SIA	42	seguir	0	4	2	I
#,S,E,G,U,I,M,O,S,$	11	seguimos	1PIA	42	seguir	0	4	4	I
#,S,E,G,U,I,S,T,E,I,S,$	2	seguisteis	2PIA	42	seguir	0	4	6	I
#,S,I,G,U,I,E,R,O,N,$	5	siguieron	3PIA	42	seguir	0	4	5	I
#,S,E,R,$	168	ser	i	43	ser	0	1	2	I
#,S,I,E,N,D,O,$	31	siendo	pE	43	ser	0	1	5	I
#,S,I,D,O,$	83	sido	pA	43	ser	0	1	3	I
#,S,O,Y,$	19	soy	1SIE	43	ser	0	1	2	I
#,E,R,E,S,$	9	eres	2SIE	43	ser	0	2	2	I
#,E,S,$	904	es	3SIE	43	ser	0	2	0	I
#,S,O,M,O,S,$	15	somos	1PIE	43	ser	0	1	4	I
#,S,O,I,S,$	3	sois	2PIE	43	ser	0	1	3	I
#,S,O,N,$	210	son	3PIE	43	ser	0	1	2	I
#,F,U,I,$	3	fui	1SIA	43	ser	0	2	1	I
#,F,U,I,S,T,E,$	2	fuiste	2SIA	43	ser	0	2	4	I
#,F,U,E,$	112	fue	3SIA	43	ser	0	2	1	I
#,F,U,I,M,O,S,$	2	fuimos	1PIA	43	ser	0	2	4	I
#,F,U,I,S,T,E,I,S,$	2	fuisteis	2PIA	43	ser	0	2	6	I
#,F,U,E,R,O,N,$	29	fueron	3PIA	43	ser	0	2	4	I
#,T,E,N,E,R,$	195	tener	i	44	tener	0	3	2	I
#,T,E,N,I,E,N,D,O,$	39	teniendo	pE	44	tener	0	3	5	I
#,T,E,N,I,D,O,$	68	tenido	pA	44	tener	0	3	3	I
#,T,E,N,G,O,$	127	tengo	1SIE	44	tener	0	4	1	I
#,T,I,E,N,E,S,$	64	tienes	2SIE	44	tener	0	4	2	I
#,T,I,E,N,E,$	440	tiene	3SIE	44	tener	0	4	1	I
#,T,E,N,E,M,O,S,$	103	tenemos	1PIE	44	tener	0	3	4	I
#,T,E,N,E,',I,S,$	10	tenéis	2PIE	44	tener	0	3	4	I
#,T,I,E,N,E,N,$	194	tienen	3PIE	44	tener	0	4	2	I
#,T,U,V,E,$	15	tuve	1SIA	44	tener	0	3	1	I
#,T,U,V,I,S,T,E,$	2	tuviste	2SIA	44	tener	0	3	4	I
#,T,U,V,O,$	53	tuvo	3SIA	44	tener	0	3	1	I
#,T,U,V,I,M,O,S,$	8	tuvimos	1PIA	44	tener	0	3	4	I
#,T,U,V,I,S,T,E,I,S,$	2	tuvisteis	2PIA	44	tener	0	3	6	I
#,T,U,V,I,E,R,O,N,$	17	tuvieron	3PIA	44	tener	0	3	5	I
#,T,R,A,B,A,J,A,R,$	53	trabajar	i	45	trabajar	0	6	2	R
#,T,R,A,B,A,J,A,N,D,O,$	28	trabajando	pE	45	trabajar	0	6	4	R
#,T,R,A,B,A,J,A,D,O,$	13	trabajado	pA	45	trabajar	0	6	3	R
#,T,R,A,B,A,J,O,$	4	trabajo	1SIE	45	trabajar	0	6	1	R
#,T,R,A,B,A,J,A,S,$	3	trabajas	2SIE	45	trabajar	0	6	2	R

#,T,R,A,B,A,J,A,$	21	trabaja	3SIE	45	trabajar	0	6	1	R
#,T,R,A,B,A,J,A,M,O,S,$	6	trabajamos	1PIE	45	trabajar	0	6	4	R
#,T,R,A,B,A,J,á,I,S,$	2	trabajáis	2PIE	45	trabajar	0	6	3	R
#,T,R,A,B,A,J,A,N,$	16	trabajan	3PIE	45	trabajar	0	6	2	R
#,T,R,A,B,A,J,E,',$	2	trabajé	1SIA	45	trabajar	0	6	2	R
#,T,R,A,B,A,J,A,S,T,E,$	2	trabajaste	2SIA	45	trabajar	0	6	4	R
#,T,R,A,B,A,J,ó,$	6	trabajó	3SIA	45	trabajar	0	6	1	R
#,T,R,A,B,A,J,A,M,O,S,$	1	trabajamos	1PIA	45	trabajar	0	6	4	R
#,T,R,A,B,A,J,A,S,T,E,I,S,$	2	trabajasteis	2PIA	45	trabajar	0	6	6	R
#,T,R,A,B,A,J,A,R,O,N,$	3	trabajaron	3PIA	45	trabajar	0	6	4	R
#,T,R,A,T,A,R,$	24	tratar	i	46	tratar	0	4	2	R
#,T,R,A,T,A,N,D,O,$	8	tratando	pE	46	tratar	0	4	4	R
#,T,R,A,T,A,D,O,$	8	tratado	pA	46	tratar	0	4	3	R
#,T,R,A,T,O,$	2	trato	1SIE	46	tratar	0	4	1	R
#,T,R,A,T,A,S,$	2	tratas	2SIE	46	tratar	0	4	2	R
#,T,R,A,T,A,$	100	trata	3SIE	46	tratar	0	4	1	R
#,T,R,A,T,A,M,O,S,$	3	tratamos	1PIE	46	tratar	0	4	4	R
#,T,R,A,T,á,I,S,$	2	tratáis	2PIE	46	tratar	0	4	3	R
#,T,R,A,T,A,N,$	9	tratan	3PIE	46	tratar	0	4	2	R
#,T,R,A,T,E,',$	2	traté	1SIA	46	tratar	0	4	2	R
#,T,R,A,T,A,S,T,E,$	2	trataste	2SIA	46	tratar	0	4	4	R
#,T,R,A,T,ó,$	5	trató	3SIA	46	tratar	0	4	1	R
#,T,R,A,T,A,M,O,S,$	1	tratamos	1PIA	46	tratar	0	4	4	R
#,T,R,A,T,A,S,T,E,I,S,$	2	tratasteis	2PIA	46	tratar	0	4	6	R
#,T,R,A,T,A,R,O,N,$	3	trataron	3PIA	46	tratar	0	4	4	R
#,U,T,I,L,I,Z,A,R,$	45	utilizar	i	47	utilizar	0	6	2	R
#,U,T,I,L,I,Z,A,N,D,O,$	21	utilizando	pE	47	utilizar	0	6	4	R
#,U,T,I,L,I,Z,A,D,O,$	19	utilizado	pA	47	utilizar	0	6	3	R
#,U,T,I,L,I,Z,O,$	4	utilizo	1SIE	47	utilizar	0	6	1	R
#,U,T,I,L,I,Z,A,S,$	2	utilizas	2SIE	47	utilizar	0	6	2	R
#,U,T,I,L,I,Z,A,$	25	utiliza	3SIE	47	utilizar	0	6	1	R
#,U,T,I,L,I,Z,A,M,O,S,$	4	utilizamos	1PIE	47	utilizar	0	6	4	R
#,U,T,I,L,I,Z,á,I,S,$	2	utilizáis	2PIE	47	utilizar	0	6	3	R
#,U,T,I,L,I,Z,A,N,$	18	utilizan	3PIE	47	utilizar	0	6	2	R
#,U,T,I,L,I,C,E,',$	2	utilicé	1SIA	47	utilizar	0	6	2	R
#,U,T,I,L,I,Z,A,S,T,E,$	2	utilizaste	2SIA	47	utilizar	0	6	4	R
#,U,T,I,L,I,Z,ó,$	5	utilizó	3SIA	47	utilizar	0	6	1	R
#,U,T,I,L,I,Z,A,M,O,S,$	1	utilizamos	1PIA	47	utilizar	0	6	4	R
#,U,T,I,L,I,Z,A,S,T,E,I,S,$	2	utilizasteis	2PIA	47	utilizar	0	6	6	R
#,U,T,I,L,I,Z,A,R,O,N,$	3	utilizaron	3PIA	47	utilizar	0	6	4	R
#,V,E,N,I,R,$	14	venir	i	48	venir	0	3	2	I
#,V,I,N,I,E,N,D,O,$	3	viniendo	pE	48	venir	0	3	5	I
#,V,E,N,I,D,O,$	15	venido	pA	48	venir	0	3	3	I
#,V,E,N,G,O,$	1	vengo	1SIE	48	venir	0	4	1	I
#,V,I,E,N,E,S,$	3	vienes	2SIE	48	venir	0	4	2	I
#,V,I,E,N,E,$	63	viene	3SIE	48	venir	0	4	1	I
#,V,E,N,I,M,O,S,$	4	venimos	1PIE	48	venir	0	3	4	I
#,V,E,N,í,S,$	2	venís	2PIE	48	venir	0	3	2	I

#,V,I,E,N,E,N,$	21	vienen	3PIE	48	venir	0	4	2	I
#,V,I,N,E,$	2	vine	1SIA	48	venir	0	3	1	I
#,V,I,N,I,S,T,E,$	2	viniste	2SIA	48	venir	0	3	4	I
#,V,I,N,O,$	9	vino	3SIA	48	venir	0	3	1	I
#,V,I,N,I,M,O,S,$	2	vinimos	1PIA	48	venir	0	3	4	I
#,V,I,N,I,S,T,E,I,S,$	2	vinisteis	2PIA	48	venir	0	3	6	I
#,V,I,N,I,E,R,O,N,$	4	vinieron	3PIA	48	venir	0	3	5	I
#,V,E,R,$	237	ver	i	49	ver	0	1	2	I
#,V,I,E,N,D,O,$	20	viendo	pE	49	ver	0	1	5	I
#,V,I,S,T,O,$	76	visto	pA	49	ver	0	1	4	I
#,V,E,O,$	39	veo	1SIE	49	ver	0	1	2	I
#,V,E,S,$	6	ves	2SIE	49	ver	0	1	2	I
#,V,E,$	43	ve	3SIE	49	ver	0	1	1	I
#,V,E,M,O,S,$	19	vemos	1PIE	49	ver	0	1	4	I
#,V,E,I,S,$	5	veis	2PIE	49	ver	0	1	3	I
#,V,E,N,$	22	ven	3PIE	49	ver	0	1	2	I
#,V,I,$	17	vi	1SIA	49	ver	0	1	1	I
#,V,I,S,T,E,$	1	viste	2SIA	49	ver	0	1	4	I
#,V,I,O,$	14	vio	3SIA	49	ver	0	1	2	I
#,V,I,M,O,S,$	6	vimos	1PIA	49	ver	0	1	4	I
#,V,I,S,T,E,I,S,$	2	visteis	2PIA	49	ver	0	1	6	I
#,V,I,E,R,O,N,$	7	vieron	3PIA	49	ver	0	1	5	I
#,V,O,L,V,E,R,$	37	volver	i	50	volver	0	4	2	I
#,V,O,L,V,I,E,N,D,O,$	6	volviendo	pE	50	volver	0	4	5	I
#,V,U,E,L,T,O,$	16	vuelto	pA	50	volver	0	4	2	I
#,V,U,E,L,V,O,$	7	vuelvo	1SIE	50	volver	0	5	1	I
#,V,U,E,L,V,E,S,$	3	vuelves	2SIE	50	volver	0	5	2	I
#,V,U,E,L,V,E,$	26	vuelve	3SIE	50	volver	0	5	1	I
#,V,O,L,V,E,M,O,S,$	5	volvemos	1PIE	50	volver	0	4	4	I
#,V,O,L,V,E,',I,S,$	2	volvéis	2PIE	50	volver	0	4	4	I
#,V,U,E,L,V,E,N,$	9	vuelven	3PIE	50	volver	0	5	2	I
#,V,O,L,V,í,$	3	volví	1SIA	50	volver	0	4	1	I
#,V,O,L,V,I,S,T,E,$	2	volviste	2SIA	50	volver	0	4	4	I
#,V,O,L,V,I,ó,$	13	volvió	3SIA	50	volver	0	4	2	I
#,V,O,L,V,I,M,O,S,$	2	volvimos	1PIA	50	volver	0	4	4	I
#,V,O,L,V,I,S,T,E,I,S,$	2	volvisteis	2PIA	50	volver	0	4	6	I
#,V,O,L,V,I,E,R,O,N,$	5	volvieron	3PIA	50	volver	0	4	5	I