UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

MÁRIO LÚCIO MESQUITA MACHADO

A Verb Learning Model Driven by Syntactic Constructions

Thesis presented in partial fulfillment of the requirements for the degree of Master in Computer Science

Profa. Dra. Aline Villavicencio
Advisor

Prof. Dr. Marco P. Idiart
Co-advisor

Porto Alegre, March of 2008.

"Infatuated by the extensiveness of the universe, it's easy to forget one of the most complex enigmas of this world: The human mind, a universe of it's own."

The Gathering - Gaya's Dream

AKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial intelligence
NLP	Natural Language Processing
POS	Part-of-Speech
RASP	Robust Accurate Statistical Parsing
UCREL	University Centre for Computer Corpus Research on Language

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Cognitive theories have been, since the second half of the last century, bringing some interesting views about language learning. The application of these theories on computational models has double benefits: in the one hand, computational implementations can be used as a form of validation of these theories; on the other hand, computational models can earn an improved performance from adopting some cognitively plausible learning strategies.

Syntactic structures are said to provide an important cue for the acquisition of verb meaning. Yet, for a particular subset of very frequent and general verbs – the so-called light verbs – there is a strong link between the syntactic structures in which they appear and their meanings. In this work, we used a computational model, to further investigate these proposals, in particular looking at the acquisition task as a mapping between an unknown verb and prototypical referents for verbal events, on the basis of the syntactic structure in which the verb appears. The experiments conducted have highlighted some requirements for a successful learning, both in terms of the levels of information available to the learner and the learning strategies adopted.

Um Modelo de Aquisição de Verbos Guiado por Construções Sintáticas

RESUMO

Desde a segunda metade do último século, as teorias cognitivas têm trazido algumas visões interessantes em relação ao aprendizado de linguagem. A aplicação destas teorias em modelos computacionais tem duplo benefício: por um lado, implementações computacionais podem ser usaas como uma forma de validação destas teorias; por outro lado, modelos computacionais podem alcançar uma performance melhorada a partir da adoção de estratégias de aprendizado cognitivamente plausíveis.

Estruturas sintáticas são ditas fornecer uma pista importante para a aquisição do significado de verbos. Ainda, para um subconjunto particular de verbos muito frequentes e gerais - os assim-chamados light verbs - há uma forte ligação entre as estruturas sintáticas nas quais eles aparecem e seus significados. Neste trabalho, empregamos um modelo computacional para investigar estas propostas, em particular, considerando a tarefa de aquisição como um mapeamento entre um verbo desconhecido e referentes prototípicos para eventos verbais, com base na estrutura sintática na qual o verbo aparece. Os experimentos conduzidos ressaltaram alguns requerimentos para um aprendizado bem-sucedido, em termos de níveis de informação disponível para o aprendiz e da estratégia de aprendizado adotada.

1 INTRODUCTION

One important and widely investigated topic for linguists, cognitive scientists, psychologists and related professionals is language acquisition. Its study is a way of shading light over issues such as concept learning, concept categorisation, social-pragmatic interactions, word meaning acquisition and other elements involved in the task as a whole, as well as in other aspects concerning human cognition.

Theories about language acquisition are also important and interesting to the field of Natural Language Processing (NLP). They can help in designing NLP systems able to successfully deal with the open-ended and dynamic nature of languages, taking advantage of features and mechanisms that enable the use of language by humans. Of particular interest for applications involving language technology is the topic of lexical acquisition, which involves the organisation of and access to the inner lexicon. This interest comes from the huge amount of words (and meanings) mastered by a normal speaker and the speed of information retrieval in such a base (JOHNSTON, 1997). Artificial Intelligence (AI) can also draw some inspiration from such research. A better understanding of the human cognitive processes may bring as side effect the improvement of current models of reasoning and knowledge representation. Likewise, for Psychology and Linguistics, the existence of computational models that investigate the requirements and conditions needed for the accuracy of a cognitive theory may influence the directions of further research and refinements of these theories.

This task of lexical acquisition – the learning of word meanings – has been investigated by researchers of all these different areas. During lexical acquisition, children map the vocalisation that they hear into inner representations (meanings), in an effort to understand communications and to communicate themselves (TOMASELLO, 2003). In order to do this, they apply as many resources as they have in hand. Of course, several types of words can be more challenging to learn than others. Verbs, for instance, seem to demand some more effort than nouns since, differently from the later, the referentials for the former are more difficult to catch. According to Tomasello: "actions and events have more fluid temporal boundaries and these are defined in different ways for different verbs" (2003, p. 47). Some authors argue that the way of learning verbs goes through the use of information provided by syntactic structures. These claims are confirmed by experiments such as those performed by Gleitman and Gillette (1995). They show that, without some syntactic information, it is very difficult to assign the meaning of a verb correctly.

The role of syntactic structures is also looked at by Goldberg (1999), in a cross-linguistic study of child language acquisition. She observed that some syntactic structures have an inherent embedded meaning, derived from general and pervasive actions, such as movement and transference, forming what she calls "constructions".

Through these structures, the meaning of a previously unknown verb, and therefore, that of a clause or sentence, can be approximated.

This is an appealing idea for NLP technologies, because structural information is relatively easy to be extracted. Thus meaning of unknown verbs can be obtained online from input data, without the need of heavy loads of information.

In this work, we investigate the influence that constructions exert in the task of meaning acquisition, either for humans or artificial learners. In our hypothesis the syntactic context can provide the basis for determining the semantics of a verb, borrowing Goldberg's (1999) idea of constructions, and the notions of syntactic influence on verb learning from Gleitman and Gillette (1995). In order to test this hypothesis we employed different computational models in a set of experiments, aiming to determine the conditions that learners would need for a successful learning process. These models receive, as input, data from the CHILDES database (MACWHINNEY, 2000) with transcriptions of child-directed speech, in an attempt to approximate the linguistic input that a child receives during language learning.

This work attempts to advance towards a better integration of findings in cognitive linguistic work and Natural Language Processing developments, for the construction of more adaptive and cognitive-based NLP technology on one side, and the empirical testing of linguistic theories through computational simulations on the other. For that we propose the implementation of a purely cognitive theory in a computational environment. This strategy allows a system to operate in a fashion more closely related to reality, leaving less space for assumptions that are not are not supported by evidence from theoretical and empirical research. The use of this theory brings, as positive side-effect the possibility of incremental learning and extraction of information from data (aided by other sources of input, of course), done in a dynamic and online fashion as the system is exposed to its linguistic environment.

This dissertation is organised as follows: chapter 2 reviews the theoretical issues underlying the work, as well as some related works. Chapter 3 presents the computational model for learning the syntactic semantic link for verbs and the experimental setup employed in this work. Chapter 4 discusses the experiments performed and results obtained, addressing questions raised along the work. Finally, in chapter 5, we discuss the conclusions and future works.

2 THEORETICAL UNDERLYINGS

In this chapter, the theoretical basis of this work is discussed in more detail. We look at the lexical acquisition by children, focusing on verb learning issues. The role of syntax in verb learning is discussed, as well as the background concepts used in this work, such as constructions, *light verbs* and prototypes. Finally, we also discuss some cognitively-based work on the mental lexicon, including a proposal made by Gaume (2005).

2.1 Word Learning

Word meaning acquisition is not a simple task. It is not just a matter of labelling things, actions, etc, but also understanding what these things really stand for and in which situations they must be referred to. Tomasello observes some of the difficulties of this enterprise:

> First, adults in many cultures do not stop what they are doing to name things for children at all. These children experience basically all words in the ongoing flow of social interaction and discourse in which adults produce many different types of words in many different types of utterances - virtually none of which present new words isolate from other words, while at the same time the adult is explicitly designating some entity with pointing or some other gesture. Second, even the most pedagogically conscious Western middle-class parents seldom play the pointing and naming game with words other than object labels; parents do not say to their children "Look! Giving" or "Look! Of". This means that the child must learn many, perhaps most, words from more complex interactive situations in which determining the adult's intended referent for some novel word is much less straightforward. Third, even in the pointing-and-naming game, things are not as simple as they first appear. When someone holds up a toy car and names it for a child, how is the child to know whether the adult is saying something like car, toy or Volkswagen? [...] Or even worse, how is the child to know that the adult is naming the object at all - as opposed to designating one of its parts or properties or its owner or some action it is about to engage in? (2003, p. 43)

Many views have been proposed for how this process occurs. As more and more details of the children's cognitive processes become unveiled, theories about acquisition become richer and able to cover a wider array of nuances, as we discuss in what follows.

2.1.1 Views on Word Meaning Acquisition

The first of these views is the *associative learning*, which is related to early theories about language learning, especially those proposed by B. F. Skinner (1957) and his

contemporaries. For the researchers of this period, language was a kind of behaviour, governed by the same principles admitted for other behaviours, like stimuli, response and reinforcement. Even in more recent days, some studies take this fashion. For instance:

> The essence of word learning is associating sounds with salient aspects of perceptual experience. In support of this view, Smith (2000) has demonstrated in several experiments that children often assume that the meaning of a novel word is the most "salient" aspect of the current non-linguistic context. (2003, p. 82)

Some experiments, however, contradicted the claims of this theory. For instance, in one experiment in which a novel word was being taught to a child, as the word was being used one object in the environment was made clearly salient while the researcher's glance was aimed at another object. When the child was led to generate the word, she used it to refer not to the most salient object, as postulated by the theory, but to that one glanced by the researcher. This finding, along with others, reinforces the thesis that the important aspect is not the salience of the object but a shared focus of attention (TOMASELLO, 2003).

In addition, this theory leaves other questions open, and does not provide a complete explanation of language learning, only covering words used to label objects. The acquisition of functional words, such as prepositions, conjunctions and so on, remains without explanation.

The second approach proposed for word learning is known as *constrained approach*. The motivations behind this view can be exemplified by Quinn's parable:

> [...] a native who utters the expression "Gavagai" and "shows" a foreigner the intended referent by pointing to a salient event as it unfolds: a rabbit running past. The problem is that since there is no shared context between interactants for this expression, there is basically no way that the foreigner can know whether the native's novel expression is being used to refer to the activity, to the rabbit, to some part of the rabbit's body, to the colour of the rabbit's fur or to any of an infinite number of things (TOMASELLO, 2005, p. 84).

This "referential uncertainty" (SISKIND, 1996) can be seen as an example of the poverty-of-stimulus[1], advocated by Chomsky (1982). He has argued that the linguistic input presented to a child was not rich enough to ensure the acquisition of all grammatical aspects of a language. Thus, the defenders of this point of view argue that the children can only learn a language by obeying some constraints that limit the search space for the word meanings. These constraints take the form of assumptions taken into account to ensure the correct acquisition.

Markman (2003), for instance, proposes two constraints: the Whole Object Constraint and the Mutual Exclusion Constraint. The former says that, in the case of a new referent, the new word must be assigned to the object as a whole, instead of some of its parts, some attribute or some action on which it is engaged. The Mutual Exclusion constraint states that, in face of objects previously named, the new word must designate some of its parts, attributes or actions (without, however, specifying how the assignments are made to each of these categories). Golinkoff, Mervis and Hirsh-Pasek

---

[1] For poverty-of-stimulus, we understand the lack of linguistic information needed to cover the grammatical issues of a language. For Chomsky, the sentences heard by a child carry not enough information to enable the learning of more abstract grammatical aspects.

(1994), on their turn, propose a set of principles, which are meant to be softer than the constraints, and acquired by the experience of the child during the learning. However, again, many words that are learned and employed early on in the learning process remain without explanation.

The third view on word learning is the *social-pragmatic*. On this view, what guides the word learning is:

> The structure of the social world into which children are born, full of scripts, routines, social games and other patterned cultural interactions and the children's social cognitive capacities for tuning into and participating in this structured social world – especially joint attention and intention-reading (with the resulting cultural learning) (2003, p.87).

The existence of periodic activities, with well-defined actors' roles, objects, actions, etc creates an environment in that is possible for the children's abilities to catch word meanings. This is done through the effort of the child in attempting to understand the communicative intentions of other persons as she interacts with them socially and linguistically (TOMASELLO, 2005). Thus, in this perspective, the child's goal is to understand what someone is trying to say, instead of finding correct mappings between word and world. According to Tomasello:

> Learning the communicative significance of an individual word consists in the child first discerning the adult's overall communicative intention in making the utterance, and then identifying the specific functional role this word is playing in the communicative intention as a whole (2003, p. 89).

In this work, our research is centred on the hypothesis of the *syntactic bootstrap*, which can be seen as a kind of constraint applied to the learning process. Thus, we are – kept the due proportions – making use of a constrained approach.

2.1.2 Difficulties in Acquiring Words

Different classes of words pose different challenges for a child during the acquisition age. The views presented in the previous section tried to provide explanations about how this learning is accomplished through an increasing comprehension of children's cognitive skills. We now take a closer look at some word categories and some of the difficulties they represent for acquisition.

Nouns seem to be the category that is easier to acquire. Indeed, research in many languages has shown that nouns are the first and more numerous words to be learned (GENTNER, 1982) (CHOI and GOPNIK, 1995). Tomasello discusses a proposal by Gentner (1982) for this phenomenon:

> In brief, her hypothesis was that the nouns children learn early in development are prototypically used to refer to concrete objects, and concrete objects are more easily individuated from their environmental surroundings than are states, actions, processes, and attributes. Concrete objects are spatially bounded entities, perceptible at a glance; whereas actions and events have more fluid temporal boundaries and these are defined in different ways for different verbs (cleaning is over when things are clean, but running and smiling have no such clearly defined endpoints) (2005, p. 47).

This relative ease in being acquired, however, applies only to common concrete nouns. But even concrete nouns have their own complexities, which are made salient by phenomena like underextention and overextention. For Barrett (1995), the underextention happens when a word is not used with its whole range of possible

referents. On the other hand, the overextension occurs when a word is used for all its referents and for some inappropriate others, which bear some perceptual or functional analogy with the true referents. More detailed explanations will be given in the section 2.2.3.

Other open classes[2] of words like adjectives and verbs are more challenging than nouns. Although some of them may have a referential in the world, they are not so easily identified as nouns. For instance, adjectives that designate attributes do not exist by themselves. The understanding of what an adjective is referred depends often on the owner of the attribute and the social-pragmatic context. The ideas of "hot", "small" and "long", for instance, are different from one object to another, in absolute terms. Yet, not all adjectives are linked to physical entities; some of them express inner states or subjective attributes. Similarly, these same problems apply to abstract nouns.

If, on the one hand, the difficulties of open class words are centred in the mapping between words and the right referent(s), on the other hand, the problem with closed-class words, like prepositions and articles, is that they have no referents to map to. The role of such words is to provide linguistic connections, rather than assigning entities.

In sum, different classes of words represent distinct sets of problems for language acquisition that require possibly specific solutions. However, a child still manages to learn all of them successfully in a short time.

2.2 Verb Learning

The last section provided a brief discussion about word learning, as well as its difficulties. Now, we are going to focus on verb learning, which poses its own challenges for the acquisition task.

2.2.1 Importance of Syntactic Cues

Verbs, like other open class words, have a referent to be mapped to in the real world. For example, *fly*, which stands for the "displacement through aerial ways". Although perceptual inputs, like vision, may help in the acquisition of verbs, they are not enough. Verbs have no spatial or temporal boundaries, like the referents of nouns and adjectives. A verb may refer to something that is going on, has already happened or is going to happen, therefore, some form of "external help" for the learning of verbs is necessary. Which exactly is the form of this external help is a topic of much discussion. For instance, Gleitman and Gillette (1995) argue that the acquisition of verbs is supported by, among other things, structural (syntactic) information. According to them: "this word-to-world pairing procedure is too weak to account fully for the mapping [between verbs and actions]. Our claim is that word learning is in general performed by pairing a sentence (syntactic object) with the observed world" (1995, p. 414).

---

[2] Nouns, verbs, adjectives and adverbs are said to be open classes. They are so considered since new words can be "invented" and added to these classes. Also, some words can belong to more than one class. Closed classes, on the other hand, are not prone to changes and additions. The set of determiners, pronouns and complementizers is stable along time and words belonging to more than one closed class are not frequent.

To reinforce their statements, they report an experiment, in which adults are watching a "no audio" videotaped interaction between a mother and her baby and are asked to guess which noun the mother speaks to the child. The goal of the experiment was to evaluate the ability of identifying referents based only on context, and it had a high rate of correct matching. In another occasion, the experiment was repeated, but this time asking the observers which verb the mother was speaking and the results show a considerably higher error rate.

Gleitman and Gillette argue that the difficulties in capturing the meaning of a verb only on the basis of context is derived from, among other things, the fact that some verbs represent non observable situations, e.g. *want, know* and *think*, as well as from a temporal disparity between the uttered verb and the action that has taken place.

The acquisition of verbs, among other constructs, grows suddenly circa 24 months of age, when the child begins to build sentences with two words, according to Lenneberg (1967), as discussed by Gleitman and Gillette:

> Perhaps an ability to comprehend the spoken sentence is a requirement for efficient verb learning. It may be that once the child has learned some nouns, she can ask not only: "What are the environmental contingencies for the use of this word?" but "What are its environmental contingencies as constrained by the structural positions in which it appears in adult speech? (1995, p. 416)

Another important point in the argument of Gleitman and Gillette (hereby G&G) is the interaction between form and meaning. Under this point of view, "the structure is the projection of meaning of the verb, that is, the surface structures are mapped from the argument structure of the verb" (1995, p. 417). Following this idea, two complementary views from the acquisition of verbs, based on the relation of between form and meaning, are presented. One of them is the *semantic bootstrap*, advocated by Pinker (PINKER, 1984), in which children can project the argument structure for verbs whose meanings were acquired by event observation, that is, a child uses the meaning as the basis for the acquisition of the syntax of a word. The other hypothesis, which is consistent with G&G's proposal, is the *syntactic bootstrap*. In this hypothesis, the syntactic structure in which a verb appears constrains the search space for the meaning of the verb. In G&G's hypothesis, children employ syntactic information as the basis for the interpretation of a new verb. This is done by means of perceptual and pragmatic impressions, paired with a linguistic event: a new verb, placed within a parsing tree built from an adult's sentence.

Gleitman and Gillette's hypothesis about the role of syntactic structures for verb learning is compatible with our work, and in fact this work can help understand the conditions under which their syntactic bootstrap can be successfully adopted for language acquisition.

2.2.2 More on verb learning: Constructions and *Light Verbs*

Goldberg (1999) also addresses the matter of verb acquisition and the link between form and meaning. She defines a set of postulates related to the categorisation and generalisation of verbs based on the argument structure in which they occur. Goldberg points out that, although a verb is not directly responsible for the form and meaning of a clause, they have some strong association between themselves. She also states that there is an embodiment of meaning in syntactic structures, in such a way that, despite the original meaning of a verb itself, when it occurs in a particular syntactic structure, it adopts another meaning: that of a construction.

Constructions, according to Goldberg (1999; 2003) are structures that pair a specific syntactic form and a particular meaning, where the meaning may not be related to particular instances of words that occur in the structure:

> Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, many constructionist approaches argue that patterns are stored even if they are fully predictable as long as they occur with sufficient frequency. (2003, p. 219-220)

Examples of constructions are presented in table 2.1, where *Subj* stands for subject of the sentence, *V* for verb, *Obj1* and *Obj2* for objects, *Obl* for oblique object and XCOMP for object complement.

Table 2.1: Examples of constructions

Construction/example	Meaning	Form
1. Intransitive Motion The fly buzzed into the room	X moves to Y	Subj V Obl
2. Transitive Pat cubed the meat.	X acts on Y	Subj V Obj
3. Resultative She kissed him unconscious.	X causes Y to become Z	Subj V Obj XCOMP
4. Double Object Pat taxed Bill the letter.	X causes Y to receive Z	Subj V Obj1 Obj2
5. Caused Motion Pat sneezed the foam off the cappuccino	X causes Y to move Z	Subj V Obj Obl

Source: GOLDBERG, 1999. p. 199

According to Goldberg the light verbs *go*, *do*, *make*, *give* and *put* embody each of these constructions, as shown in table 2.2.

The importance of these structures lies in the fact that their inherent meaning supplies cues, for example, to the meaning of unknown verbs appearing in them, functioning as an initial step for generalisation.

However, if the learning of a verb is aided by these constructions, the question then is where do these constructions come from? Goldberg argues that, initially, the acquisition of such structures follows a verb-to-verb basis, that is, the child associates the argument structures individually for each verb. However, as Goldberg points out, this cannot continue indefinitely, due to the number of verbs in a language. Thus, children must have a way to make generalisations over particular instances previously learnt. According to Goldberg: "semantically similar verbs show a strong tendency to appear in the same argument structure constructions"(1999, p. 200). As for the construction's inherent meaning, Goldberg suggests that the generalisation of the constructional meaning is based on the meanings of the so-called *light verbs*, verbs like *go*, *do*, *make*, *give* and *put*, highly frequent and with very general meanings. For instance, *go* can be used in the same places of more specialized verbs indicating motion, such as *walk*, *drive*, *swim*, *fly*, etc. Table 2.2 shows some examples of light verbs, as well as the constructions related to them.

Table 2.2: Some examples of light verbs along with respective constructions

Light verb	Meaning	Construction
Go	X moves Y	Intransitive Construction
Do	X acts on Y	Transitive Construction
Make	X causes Y to become Z	Resultative Construction
Give	X causes Y to receive Z	Ditransitive Construction
Put	X causes Y to move Z	Caused Motion Construction

Source: GOLDBERG, 1999. p. 202

In this work we investigate three of these constructions, focusing on the light verbs *go*, *put* and *give*.

There seems to be a strong correlation between their high frequency and their very general and polysemic semantics, because lexical items with more general meanings are "applicable" in a wider range of situations. Furthermore, she says that light verbs codify meanings that are highly relevant to the daily human experience: action scenes, movement, cause, transferring and so on.

According to Goldberg:

> The fact that children learn the light verbs so early and use them so frequently may play a direct role in the acquisition of argument structure constructions in the following way. Children are likely to record a correlation between a certain formal pattern and the meaning of the particular verb(s) used most early and frequently in that pattern. This meaning would come to be associated with the pattern even when the particular verbs themselves do not appear. Because light verbs are most frequent than other verbs and are also learned early, these verbs tend to be the ones around which constructional meaning centres (1999, p. 208).

Just like Gleitman and this work. Their theories provide the connection between the syntactic patterns and the meanings, as well as explanations about how this meaning is constructed.

2.2.3 Prototype Theory

The concept of *light verbs,* as adopted in this work, can be seen as defining a prototypical referent for the actions denoted by a construction and the unknown verbs occurring in it. Therefore, we now discuss some relevant aspects of a related work, Barrett's Prototype Theory (BARRETT, 1995). This is an influential theory in the study of children's acquisition of referents, according to which the meaning of a word is initially acquired in the form of prototypical referents for this word. For instance, a particular "car" may represent the referent for 4-wheeled vehicles. Barrett says "This prototypical referent effectively functions for the child as a specification of the clearest and most typical referent of that word" (1995, p. 378). Furthermore, he says:

> The child then generalizes the word to other referents on the basis that they share common features with this prototype. Referents which have many features in common with the prototype are highly typical referents of the word, while referents which have relatively few features in common with the prototype are atypical or peripheral referents of the word (1995, p. 378).

For instance, "cat" may be considered a prototypical referent for felines. An explanation of what would be a tiger, a cougar, etc, would be done on the basis of a cat – a tiger is a big cat, with black and dark yellow stripes. That happens because tigers

and cats share common features, such as pawns format, long tail, some (similar) habits and so on.

Barrett asserts that this theory readily sustains the finding that referents that are included in the extension of a word are not always linked by an invariant set of common features. That is, all referents must share some of the same common features with the prototypical referent, but not necessarily the same features with each other. What makes tigers and jaguars similar between themselves is what they have in common with cats, their central referent.

This theory also supplies an explanation for the fact that, when children underextend some words - that is, when they do not use them with all their referents - they usually exclude the more peripheral referents (that is, those ones with less features in common with the prototype) from the extension of the word, and not the more typical ones (BARRETT, 1995). For instance, if the prototype of cat for a child were his mother's cat, an extreme example would be the use of "cat" only to assign cats with the same colour, disregarding cats of different colours.

The Prototype Theory was employed to explain the comprehension of words for younger children:

> When children are tested for their comprehension of words, which they spontaneously overextend in production by asking them to "show me X" or "give me X" (where X is the overextended word), they usually begin by choosing a central typical referent rather than an atypical overextended referent (1995, p. 379).

One of the problems with this theory, according to Barrett, is the lack of consensus about what exactly is meant by a prototype. Some researchers argue that a prototype is "a generalized abstract schema which represents the central tendency of a set of specific referential instances that have been experienced by the child" (1995, p. 379). Barrett also states that some researchers argue that prototypes are "holistic mental representations of individual referential exemplars that the child encountered" (1995, p. 379). Barrett, on his own, holds that:

> although prototypes may first be acquired as mental representations of individual referential exemplars, these representations are subsequently analyzed by the child into their constituent features, with the result that these prototypes eventually come to consist of correlational clusters of perceptual and functional features (1995, p. 379).

Moreover, Barrett argues that this theory does not explain the acquisition of context-bounded words, social-pragmatic words or the process of de-contextualization, and are missing theoretical constructs to explain how the overextension is rescinded.

In our context, the light verbs can be seen as prototypes for more specific actions. Making a parallel with this theory, what make verbs like *run*, *swim* and *fly* similar to each other is what they have in common with their prototype: *go.* This idea holds for the other light verbs and specific verbs, as well. This theory comes as a cognitive support for our investigation.

2.3 Mental Lexicon

Along this chapter, we have briefly discussed some cognitive and psycholinguistic views on language acquisition, as well as some of the difficulties associated with the

task, especially in relation to verbs. To finish the discussion, we now focus on the "place" where words and meanings get together: the mental lexicon.

For the purposes of this discussion, mental lexicon is defined as the repository of words and their respective meanings, along with other information, such as spelling, syntactic features, phonemes and so on. There is not, yet, an agreement among the researchers about what really is, and how it is organised. Elman (2004) raises some concerns about it:

> What does the lexicon actually look like? The metaphor of lexicon-as-dictionary is inviting, and has led to what Pustejovsky has called the 'sense enumeration model'. In that view, a lexical entry is a list of information. Just what information actually goes into the lexicon is a matter of some debate, although most linguists agree that lexical entries contain information regarding a word's semantic, syntactic and phonological properties. Some accounts of language processing have argued that lexical entries contain very fine-grained information about, for example, grammatical usage. The common thread in the vast majority of linguistic theories is to see the lexicon as a type of passive data structure that resides in long-term memory (2004, p. 301).

In this work, we investigate the acquisition of the connections between words and meanings in the mental lexicon and among words themselves, presenting an alternative view on how this structure can emerge from the linguistic environment. Deeper considerations into this topic are out of the scope of this work.

One fact that calls our attention over the lexicon is its fast accessing response. According to Johnston (1997): "native speakers can recognise a word of their language in 200 ms or less and can reject a non-word sound sequence in about half a second". Taking into account the number of words mastered for a native speaker, these numbers seem expressive. Further, she says:

> This speed is astonishing given how many words you would have to search through if you systematically examined the contents of your mental lexicon in order to reject such non-words. Just how many words are we talking about? (1997, [s. n.])

Johnston argues, also, that highly frequent, highly familiar words are accessed faster than less commonly used words. In an ordinary dictionary, however, these retrieval times are more or less equal. One possible hypothesis for this fact is the storage of these words in many different places of the brain. Also homographs are found faster then non homographs, implying that they are multiply represented for the variety of their meanings (JOHNSTON, 1997). She further relates some experiments on whose basis she argues that morphologically related words are stored together in the mental lexicon, much like a real lexicon – which seems to suggest the existence of a kind of connection among them. Moreover, for the way that words are stored, she says:

> Words are not stored in separate compartments in the mind; they coexist in an elaborate network of associations. When a word is used the activation in the mental lexicon spreads over this network of associations. Words are not only associated with meanings. They are associated with each other (JOHNSTON, 1997, [s. n.]).

This point of view is compatible with that of Elman's, for whom, words are for the brain the same as any other stimulus (2004).

Johnston also discusses some models for choosing and representing words. She says that these models must deal with the following aspects of words: the lemma (which she

defines as a combination of abstract meaning and word class - the idea behind a word) and the sound sequence (its sequence of phonemes).

The first model is featured as a sequence, in which, meanings are grouped apart from sounds, with a marked pathway from the former to the later. Also, words belonging to the same category (family, using her words) may be activated simultaneously, and this can be a possible cause for some incorrect choices of words (when instead of using the target word, the speaker uses another word belonging to the same meaning family). Some other errors, when a phonetically similar word is used instead of the target, have a similar explanation, that is, the wrong choice among all the activated similar sound sequences. This model, however, fails to explain choice errors that combine meaning and sound mistakes, as for example, changing "beaver" for "teaser". The following picture depicts graphically what this model would look like. A diagram for this model can be seen on Figure 2.1.
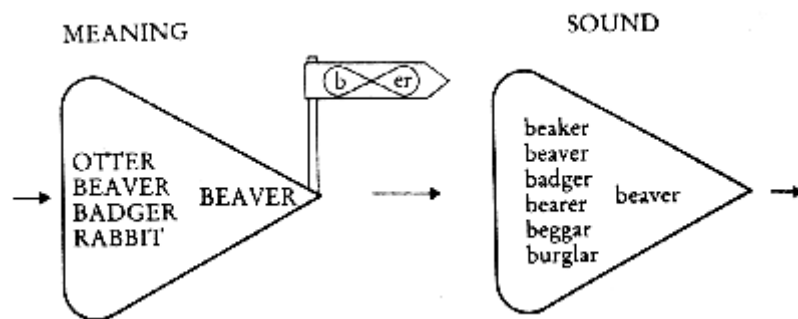
Figure 2.1: A stepping-stone model (JOHNSTON, 1997, [s. n.]).

She names the second model as a "waterfall" model. In her words:

> The waterfall model allows a person still to be thinking about the meaning as they select the sound. All the information activated at the first stage is still available at the next stage, cascading down to the next trough of water on the hillside. This model suggests that word selection is not just a case of following one word through from beginning to end, but is often a case of controlling and narrowing down a cascade of possible words. The error 'badger' for 'beaver' is explained by assuming that the semantic information about small animals was cascading down as the outline phonology was picked. The difficulty of this model is that waterfalls cannot flow backwards. Just as selecting the phonology requires semantic evidence, so phonology may be needed to narrow down the semantics. For example, consider the case of searching for members of a particular category. If I say 'think up the names of some woodland animals', you may get stuck after rabbit and squirrel. If I then prompt you with "Beginning with b" you might suddenly produce 'beaver' and 'badger'. Sounds can activate meanings, and meanings can activate sounds (1997, [s. n.]).

Figure 2.2 presents a diagram depicting the waterfall model:
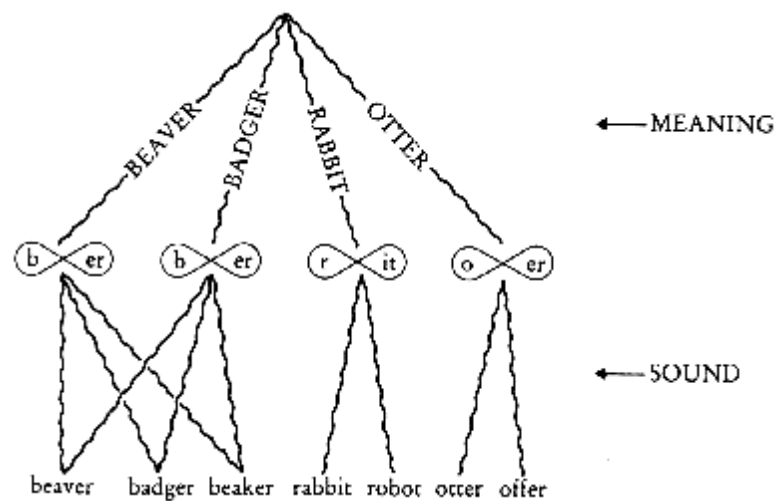
Figure 2.2: A waterfall model (JOHNSTON, 1997, [s. n.]).

In the third model, the meanings are connected among themselves, as well as the phonologic information, and these two are linked via bidirectional connections. Thus, a cognitive (or phonetic, as well) stimulus spans out through the connections, activating the words on its way.

> [...] those that are relevant get more and more excited, while those that are unwanted fade away [...]. Since the current is flowing to and fro, anything which is particularly strongly activated in the semantics will cause extra activation in the phonology, and vice versa. (1997, [s. n.])

As this model allows more then one word being activated at the same time, it is possible to utter a wrong word if it has phonology close to the right one.
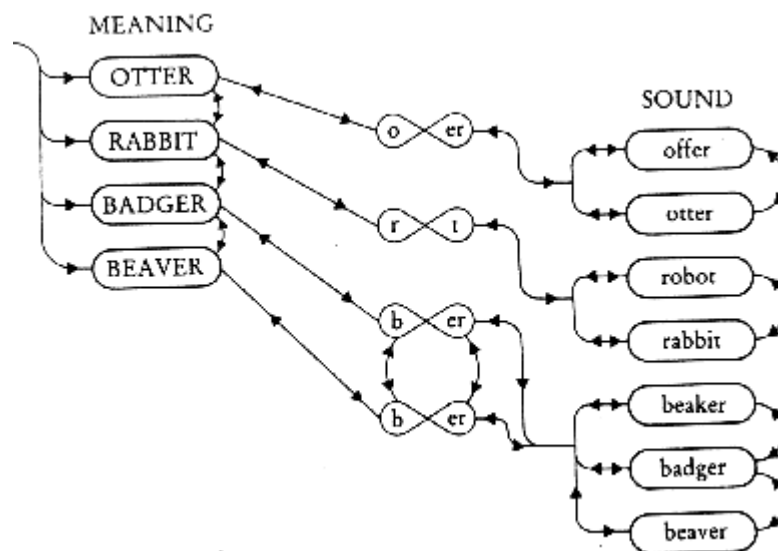
Figure 2.3: An interactive activation model (JOHNSTON, 1997, [s. n.]).

Johnston has offered us a view of how the words are connected with meanings and sound representations. Now, let's take a look on the approaches for meaning representation.

The proposal of Jackendoff (JACKENDOFF, 1975) goes towards a semantic representation in a formal and logical way. He builds his opinions upon Katz-Postal Hypothesis (hereby K&P), who suggests that semantic interpretation is performed by a set of projection rules. These rules add to the semantic representation of a sentence those parts of its content and organization not due to the lexical items, that is, the part of the interpretation traceable to the syntactic structure itself.

The syntactic structure of a sentence is generated by the application of a sequence of rules. The differences in structure between two sentences are produced by differences in the sequence of the rules that generate the sentences. Since two sentences containing the same lexical items can differ in meaning only if different sequences of rules have applied, it seemed natural to Katz and Fodor to attribute the meaning contributed by the structure to the operations of the rules themselves. That is, for each rule in the grammar, there would be an associated projection rule telling how the former contributes to the meaning of sentences to which it is applied.

Wierzbicka (2004) examines the lexicon on the basis of the semantic-conceptual structure – that is, the body that composes the meaning of the words – beneath the lexicalizations. This structure is built around what she calls "semantic primes", which are concepts for which every language has a lexicalization, for example *feel*, *die* and *not*, according to the author (in her paper there is a list of such words). According to Wierzbicka, there are as many as sixty of these semantic primes. The primes themselves are also used to compose the explanations of other concepts of higher level. These explanations can be thought of as the "meanings" of words. For instance, the meaning of "to envy someone" can be composed in the following way:

(X envies Y)
X thinks like this about Y
  "something good happened to this person
  it didn't happen to me
  this is bad"
when X thinks like this, X feels something bad because of this
  like people feel when they think like this (2004, p. 3)

According to Wierzbicka, these primes are the core around which the other meanings can be constructed and through which can be understood. Taking as basis the Polish lexicon, the author says that hundreds of words have their meanings composed directly from the primes.

Semantic molecules, which are lexicalizations built directly upon the primes, are found in a higher level of complexity. These molecules behave as if they were single units, taking part in the explanations of other concepts. For instance, "hand" can be explained in terms of primes and, in its turn takes part on the meaning of other words, like "towel", "cup", "catch" and so on.

Most of the words are composed by different combinations of primes and/or molecules. Wierzbicka postulates some templates of explanation, employed by some categories, in such a way that the members of a category make use of the same template, only with different fulfilment. For instance, Wierzbicka explains what is a mouse (the example is quite big, so, we chose to cut some parts.

a kind of small animal

they live in or near places where people live
sometimes there can be many small animals of this kind in one place
they are very small

a person can hold one easily in one hand
not many people want to hold them

they don't want to be near people or other animals
when people or other animals are near they make no noise
they hide from people and animals in places where people and animals can't reach them
they move in places where people live looking for something to eat when people are not there
they can move very quickly
they can move without making any noise [...] (2004, p. 9-10)

Her proposal, despite providing good insights and suitable explanations, does not clarify how these explanations are represented. Therefore, it is too abstract for the elaboration of (computational) models based upon them.

Elman (2004), on his turn, proposes a different view of the lexicon. Much like any other stimuli, he says that words are a kind of stimulus, as well. Moreover, according to him, words are not located in a particular place. Instead, they are spread over the connections of the neurons. It is worth noticing that his approach, differently from the other ones, is in terms of a neuronal representation (instead of a cognitive and more abstract model). Using the Simple Recurrent Network (ELMAN, 1990) he performs some experiments to validate his claims, achieving a certain degree of success, in that the network has built some internal representations, in such a way that the words kept some hierarchical relationship (ELMAN, 1990; ELMAN, 2004).

2.4 Lexical Resources

So far we have looked at the acquisition and organization of the human lexicon. However, for many years much work on NLP and Computational Linguistics has been devoted to the construction of machine-readable electronic lexical resources such as dictionaries, thesaurus and ontologies. Some of these resources are organized in such a way that their component words are connected by semantic relations (synonymy, hyperonymy, meronymy, etc), and we now discuss the relevant ones for this research. Since in this work, we investigate how to establish the connections among more specific verbs and more general ones (the light verbs), it is useful to have a standard for comparisons.

2.4.1 Wordnet

Wordnet is a lexical database, formerly only of the English language, but in recent years extended to several other languages like Italian, Portuguese and others. Wordnet was created under the direction of psychology professor George A. Miller, according to whom "its design is inspired by current psycholinguistic and computational theories of human lexical memory" (1993, p.2). Further, he says: "The most ambitious feature of Wordnet, however, is its attempt to organize lexical information in terms of word meanings, rather than word forms" (1993, p.3).

Wordnet has begun as a means to perform conceptual searches in a dictionary, instead of merely alphabetical ones (MILLER et al, 1993, p.2). The construction of

Wordnet came from psycholinguistic studies on how words are connected, inside the human cognition. About it, Miller says that:

> [...] linguists became increasingly explicit about the information a lexicon must contain in order for the phonological, syntactic, and lexical components to work together in the everyday production and comprehension of linguistic messages, [...] Beginning with word association studies at the turn of the century and continuing down to the sophisticated experimental tasks of the past twenty years, psycholinguists have discovered many synchronic properties of the mental lexicon that can be exploited in lexicography. (1993, p.2)

On Wordnet, nouns, verbs, adjectives and adverbs are organized in synonym sets, each representing one underlying lexicalized concept. Also, these sets, known as synsets, are connected through some semantic relations, for each part of speech:

- Nouns
  - hyperonyms: Y is a hyperonym of X if every X is a (kind of) Y (*person* is an hyperonym of *girl*)
  - hyponyms: Y is a hyponym of X if every Y is a (kind of) X (*girl* is an hyponym of *person*)
  - coordinate terms: Y is a coordinate term of X if X and Y share a hyperonym (*car* and *truck* are coordinate terms because they share the hyperonym *vehicle*)
  - holonym: Y is a holonym of X if X is a part of Y (*car* is an holonym for *wheel*)
  - meronym: Y is a meronym of X if Y is a part of X (*wheel* is a meronym for *car*)

- Verbs
  - hyperonym: the verb Y is a hyperonym of the verb X if the activity X is a (kind of) Y (e.g. travel to movement)
  - troponym: the verb Y is a troponym of the verb X if the activity Y is doing X in some manner (e.g. *lisp* to *talk*)
  - entailment: the verb Y is entailed by X if by doing X you must be doing Y (e.g. *sleeping* by *snoring*)
  - coordinate terms: those verbs sharing a common hyperonym (*walk* and *fly* to *displace*)

- Adjectives
  - related nouns (*beauty* to *beautiful*)
  - participle of verb

- Adverbs
  - root adjectives (*reckless* to *recklessly*)

Wordnet has found great application in NLP, being used in tasks such as word sense disambiguation, concept identification, as well as applications that need information about the semantic relationship among words.

2.4.2 Mental Lexicon

The Mental Lexicon (GAUME, 2005) is originated from the studies of Gaume and his staff about the lexical relation *analogy*, which he defines more formally as inter-domain co-hyponymy. For them, analogy is a kind of "semantic approximation", a hypothesis that can be used to explain some apparent mistakes made by children in their verb choices, as saying that a book is *broken*, instead of *torn*. In order to study the analogy in a computation al way on a verb lexicon, Gaume and his associates created the Mental Lexicon.

The Mental Lexicon is a graph structure extracted from dictionaries. On the perspective of Gaume "dictionary definitions carry meaning, they do so by the network they establish between the words which are the entries". On the construction of this graph, he assumes that the dictionary entries are nodes and "admitting the existence of an edge of a node A to a node B if and only if entry B appears in the definition of entry A". Also, Gaume and his staff have created an algorithm – PROX – to compute the distance between two nodes.

Over the experiments, it was shown that closer verbs were hyperonym or synonym from each other, while less close verbs has a relation of metaphor. (GAUME, 2005). Also, in his words:

> When distance is short, the terms are bounded by a relation of intra-domain analogy, which connects terms from the same semantic domain; when distance is the distance is a little greater, the terms are bound by a relation of "inter-domain" analogy which relates two terms from different semantic domains. (2005, p. 8)

As we could see, this is a valuable tool, either as a research workbench, either as a way of validating experiments about word learning/application.

2.5 Ontology Extraction from Corpus

From a less cognitive and more technological perspective, our work can be seen as the extraction of a verb ontology from corpora. On this context, ontology is a sort of entities holding some relation among them (usually hierarchy and meronymy). Under this prism, our work attempts to build a lexical ontology, but just with hierarchical relations. In this section, we are going to overlook other works related to ontology extraction from corpus.

Reinberger and Daelemans (2004) present a study, on which they investigated statistical approaches to evaluate verb-object relationships, in order to build a basis of semantically related words and establish semantic links between them, using machine learning strategies. According to them:

> We rely on the principle of selectional restrictions that states that syntactic structures provide relevant information about semantic content in that case, that heads of object phrases co-occurring with the same verb share a semantic feature" (2004, p. 41).

Differently from the present work, they intended to extract nominal relations, instead of verbal ones. Also, they employed a domain-specific corpus of medical abstracts related to hepatitis disease.

Faure and Nédellec (2007), on their work, present a machine learning device - ASIUM - which is aimed to extract subcategorization frames and verbs from a parsed corpus. According to the authors:

> The inputs of ASIUM result from syntactic parsing of texts. They are subcategorization examples and basic clusters formed by headwords that occur with the same verb after the same preposition (or with the same syntactical role). ASIUM successively aggregates the clusters to form new concepts in the form of a generality graph that represents the ontology of the domain. Subcategorization frames are learned in parallel, so that as concepts are formed, they fill restrictions of selection in the subcategorization frames. ASIUM method is based on conceptual clustering (2007, p. 26).

The extraction of the subcategorization frames and the ontology is completely unsupervised. However, a human expert must name the concepts generated for the ontology. Differently from the previous work, which aimed to a lexical ontology, the target of this one was a domain ontology. It employs, also, domain-specific texts for this task

Herbelot (HERBELOT, 2007) performed an extraction of conceptual clusters from corpus through the use of user-defined seeds. She employs the cluster extraction as a way of extracting an ontology (hierarchical relations). She claims that main advantage of the clustering method is that "it allows users to find hyponymic relations that are not explicitly mentioned in the corpus" (HERBELOT, 2007). In order to accomplish the task, she uses the distributional similarity hypothesis, as the previous work. The contexts are composed by a triple, in which the central word is a seed. Any other word that appears surrounded by the same words is considered to be similar to the seed. After filtering out the weaker instances and calculating the reliability of the reminder ones, these are used as seeds and the process is repeated. At the end of the process, clusters of semantically related words are formed around the seeds

As these works made clear, there are many techniques for extracting relations from text, with varying degrees in complexity and cognitive inspiration.

3 THE EXPERIMENTAL SETUP

3.1 Preliminaries

In the previous chapters, the theoretical foundations of this work were discussed in a deeper degree, in order to make clear on which theories this work is based upon. We also presented state-of-the-art of some closely related topics. In this section, we are going to show how these concepts are adopted in this work. The hypothesis that we investigate in this work is that it is possible to find an approximation for the meaning of a verb on the basis of its syntactic environment and for testing whether it holds and under what conditions it is successful, we perform some experiments which are described in the next sections. The design of the learning systems, the methodology and other implementation issues will be detailed. Figure 3.1 shows the architecture proposed.
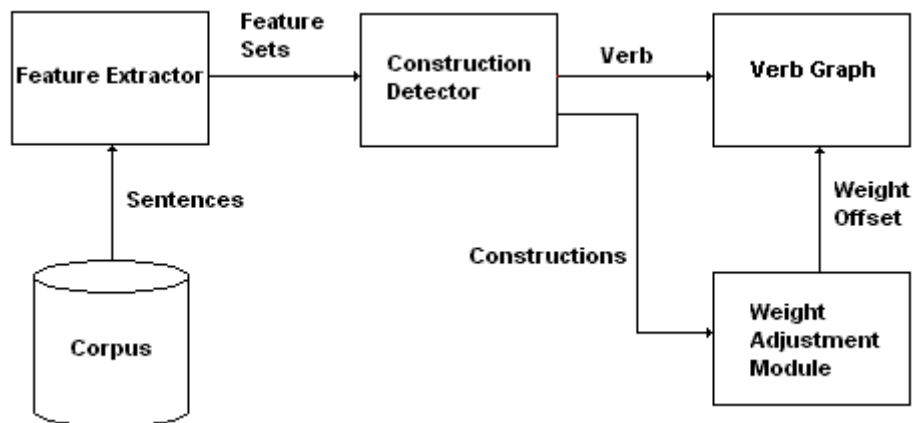
Figure 3.1: Block diagram depicting the connection among the modules of the work

First of all the input data is pre-processed by a feature extractor, which is responsible for collecting the syntactic features from each sentence, in order to subsequently identify the constructions present on it. Part of this data is to be later used as gold standard against which the results of the learners will be evaluated. The construction detector uses the features collected in the previous stage to identify a construction. The verb graph is the structure that keeps the knowledge acquired by the model. It comprises a set of nodes, representing the verbs, connected by weighted edges. The weight of each edge is adjusted in order to reflect the degree of evidence about the semantic relationship between the connected verbs. More information about each stage is given in the sections below.

3.2 Data

3.2.1 Raw Data

In order to provide a more naturalistic setting for the learners, resembling, to some extent, the linguistic environment in which a child is exposed, we use data from the CHILDES database as input to the learners.

In this section we describe the database and the corpora selected as well as the pre-processing performed.

The raw material employed for the experiments along the development of this work was obtained on the CHILDES Database (MACWHINNEY, 2000). CHILDES system (Child Language Data Exchange System) consists on a set of databases and tools for the study of conversational interactions: "These tools include a transcription database, programs for computational analysis of the transcriptions, methods for linguistic coding and systems for connection of transcriptions to digitalized audio and video".

The databases consist of, mainly, transcriptions of recorded interactions among adults and children. The goal of this project is to make data available for researches about language acquisition by children.

The Databases of CHILDES cover transcriptions of talks in a wide array of languages. Besides language acquisition, there are transcriptions that are suitable for the study of second language acquisition, adult's aphasia, language impairments, among other issues.

The portion of CHILDES employed on this work consists of longitudinal English corpora covering experiments with children of many ages, in different situations and contexts (school, family, etc). From the set of corpora contained on CHILDES, we chose three of them for the initial experiments of hypothesis testing/validation: Bates, Sachs and Brown.

These corpora were manually analyzed for obtaining a distribution of use of the light verbs considered (*go*, *put* and *give*). The distributions were consistent with those reported by Goldberg (1999). Figure 3.2 shows the distribution of the 25 most frequent verbs on the Bates corpus.
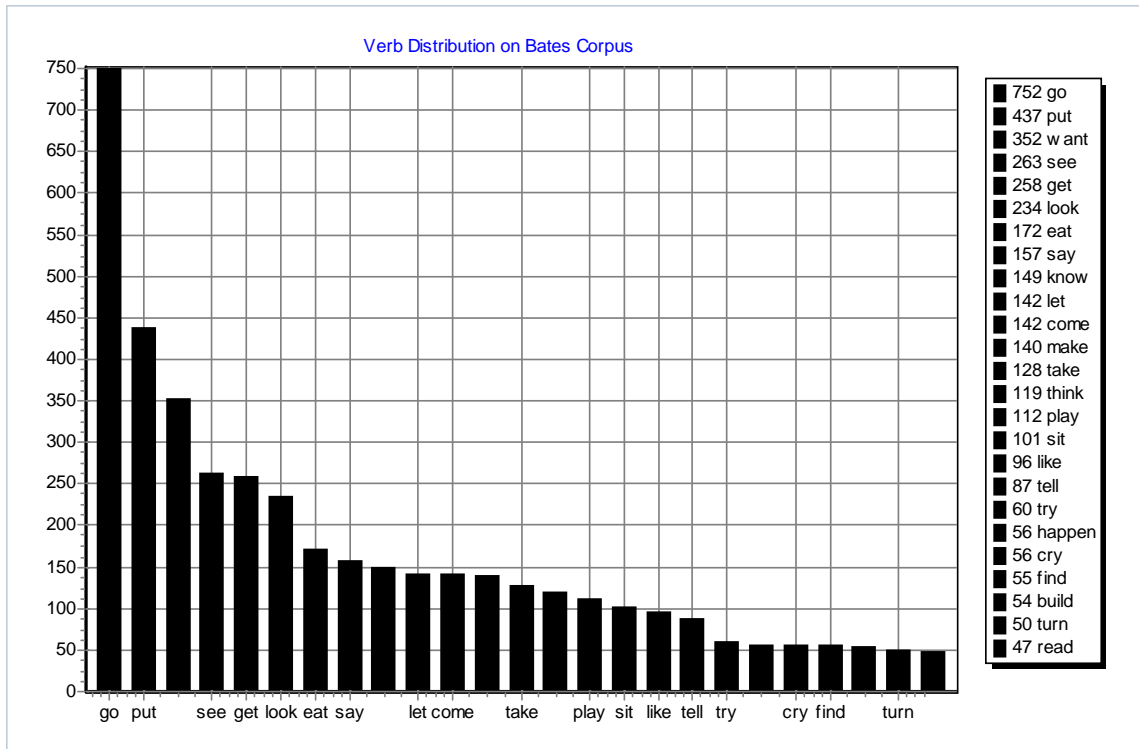
Figure 3.2: Frequency distribution showing the 25 most frequent verbs on the Bates corpus

Confirming Goldberg's analysis, we can see two of the light verbs studied here as the most frequent ones (*go* and *put*). For the verb *go*, its occurrences were analyzed individually, yielding the Chart on figure 3.3.
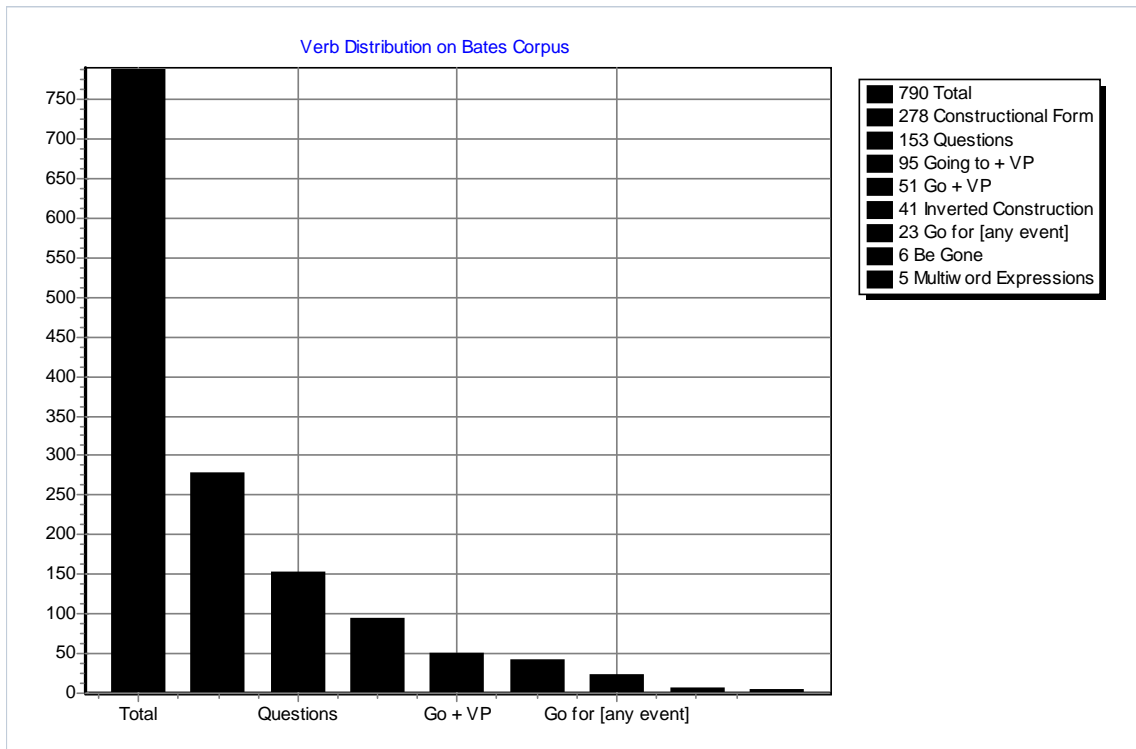
Figure 3.3: Frequency of usages of the *go* verb.

Figure 3.3 shows the frequency of different types of *go* uses. As can be seen, the most frequent is the form associated with moving actions (as pointed out by Goldberg). The items "Inverted Constructions" and "Question" showed the same elements of the constructional form, although in a different order, but with the same overall meaning of moving.

For the experiments, we extracted only adult-uttered sentences from the Brown, Sachs and Bates corpora. These corpora were stored in a POSTGRE-SQL database. It contains along with indexing and localization information:

- a table with each of the sentences;

- a table with the parsed version of each sentence, containing the derivation tree for the sentence;

- a table with the tagged version of each sentences, with all the words in them annotated with part-of-speech (POS) tags; and

- Other tables, containing further processing of the sentences, which are not used in this work.

The sentences were parsed using a robust parsing system, RASP (BRISCOE and CARROL, 2002), and the POS-tags used in the annotation are from the CLAWS2 tagset (GARSIDE, 1987), as described by Buttery and Korhonen (2005). Different levels of the available information about these sentences were used in the distinct tasks and experiments described in this work, as will be explicitly informed whenever relevant.

We employed all the sentences of these corpora, except for those ones index mismatch. It is worth noticing the difference in the amounts of occurrences in both figures, although they feature the same corpus. It happens due to a mismatch (for some indices) among the tables. For the same index value, there is a difference between the parsed and tagged sentences. The chart on figure 3.3 was built on the basis only of the tagged sentences. The chart on figure 3.2, on the other hand, was constructed during the feature extraction (which demanded a match between the tagged and parsed sentences). We used 12466 sentences from the Bates corpus, 10805 from the Sachs corpus and 11463 sentences from the Brown corpus.

3.2.2 Feature Extraction

As described in section 2.2.2 each light verb is related to a specific syntactic and semantic structure. In this work we focus on three in particular, whose structures are shown in table 3.1.

Table 3.1: Light verbs and constructions researched in this work

Verb	Syntactic Pattern	Meaning	Example
go	Subj V Obl	Movement	Sarah goes to the beach.
put	Subj V Obj Obl	Caused Movement	Joe puts his gloves inside the box.
give	Subj V Obj1 Obj2	Transference	Lisa gave flowers to her mother.

In this table we show the syntactic structure in the second column, the semantic in the third and an example in the fourth. For this work we adopt Goldberg's definitions for these constructions. In addition we further distinguish an alternative syntactic pattern. These finer-grained distinctions are made in some of the experiments since in this work we attempt to determine which type of information and how much of it is necessary in order to successfully distinguish each of these constructions, and correctly classify new instances. For doing that, we concentrate on the syntactic complements of the verb, with the exception of subjects, since the former seem to provide better means for differentiating the constructions than the latter.

- <dest> stands for a destination. Objects of this type, usually, are PPs, being the head noun a kind of place and the preposition conveying some meaning of direction, like "to", "towards", "through", etc. Also, <dest> can be a determiner (here, there, etc).

- <something, somebody> denotes objects or people, being represented by a NP or a pronoun (which is considered a person, except in the case of "it").

- <place> follows a locative pronoun or PP rule similar to <dest>, but in this case the preposition is more related to a location ("in", "on", "under", "above") instead of destination. (Quoted above).

- <recipient> is also a PP, but in this case the preposition (to) is only a (semantically empty) case marker, with the <recipient> following the same rule as <dest>. However, its head noun stands for a person or organization, rather than a place.

Thus, for each VP, the following set of features was collected:

- type of object (for each object): PP or NP;

- specific preposition (in case of a PP) or null value – "?" – otherwise;

- head noun of the NP phrase (or embedded NP phrase in the case of PPs) and its POS tag;

Each sentence would result in one or more of these sets of features, in the form of a feature vector. For instance, the sentence below yields the feature vector described in table 3.2.

```
are you going to give some orange to the piggie ?
```

As can be seen, give has two complements: an NP (*some orange*), whose head is *orange*, having the tag NN1 (single concrete noun). The semantic category for the head (which will be better explained in the section 4.1.4) is *food*. The second complement is a PP, having "*to*" as its preposition, *piggie* as its head, with the tag NN1. For this word, no hyperonym was found in Wordnet.

Table 3.2: Feature vector in details

Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.
NP	?	NN1	Food	PP	to	NN1	?	?	?	?	?

The feature vector shown on the table 3.2 is related to the verb *give*. In every sentence, each VP yields one feature vector (except for the verb *to be*, which trivially is not related to the light verbs considered here – or any other). So, in the sentence above, we would have two feature vectors: one for the verb *give* (table 3.2) and other to the verb *go*, which can be seen on the table 3.3.

```
VP , ? , VV0 , ? , ? , ? , ? , ? , ? , ? , ? , ?
```

Table 3.3: Feature vector for the verb *go* in the example.

Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.
VP	?	VV0	?	?	?	?	?	?	?	?	?

In the examples on the tables 3.2 and 3.3, the "?" appearing on the vectors stand for missing elements. On the example, we have no preposition on the first phrase.

This process of automatic feature extraction does not have 100% recall and precision, due to errors in the tagging and parsing of the input sentences. They cause (a) many potentially relevant sentences to be left out because they do not match one of the extraction patterns, while (b) conversely selecting sentences that are not relevant, but that match a pattern because of an error in parsing or tagging. For example, common tagging errors were those of incorrectly tagging nouns as verbs and conjunctions as prepositions.

Give it a little push

Table 3.4: Feature vector for the verb *give* in the example.

Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.
?	?	PPH1	?	PP	A	NN1	object	?	?	?	?

On the sentence above, "a" is returned as a preposition.

A frequent parsing problem found during the feature extraction was that of the PP-attachment, which is widely discussed in the NLP literature (MANNING and SCHUTZE, 1999). For this work meant that several sentences were extracted as a potentially relevant where a PP modifying a noun was incorrectly parsed as a complement of the verb. Roughly speaking, a PP-attachment error happens when a parser assigns a PP as object of a verb, when it really is a complement of a noun, or vice-versa. This is particularly problematic to this work, because it creates false positives (an object that the verb does not really have) or hides true features. For instance:

Want to put this one in the chair?

Table 3.5: Feature vector for the verb *want* in the example.

Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.
VP	?	VV0	?	PP	in	NN1	object	?	?	?	?

After this stage we have a file containing one or more feature vectors for each sentence. The next section will present some classifiers that used these vectors to identify to which constructions these sentences belonged.

3.3 The Learners

The assignment of a construction to a verb, and the further matching with their correspondent light verbs, is not trivial. In strictly syntactic terms, different constructions may share features, leading to a problem of ambiguity for the learner, and identification based uniquely on syntactic elements would lead to mistakes. For instance, both *put* and *give* take use of two objects, that may be very similar, sometimes, but they have different meanings. Therefore, we aim to determine the influence of ambiguity in this task, and test a machine learning algorithm and its requirements for success. For our study we chose decision trees (QUINLAN, 1996). In order to determine their suitability for this task we evaluate them using a number of different experimental conditions, starting from a linguistic environment where they have access to purely syntactic elements for the decision, to one where they can access some semantic information. In this section we are going to introduce decision trees. The experiments will be presented in the next chapter.

3.3.1 Decision Trees

Decision trees (QUINLAN, 1996) are a kind of classification device, which model a decision-taking process for classifying a set of inputs. The decision process comes from the sequential comparison of attributes with some constant values. A decision tree algorithm models the acquired information as a tree structure in which each branch represents a choice between a number of alternatives, and each leaf node represents a classification or decision. According to Witten and Frank (2005):

> Nodes in a decision tree involve testing a particular attribute. Usually, the test at a node compares an attribute value with a constant. However, some trees compare two attributes with each other, or use some function in one or more attributes. Leaf nodes give a classification that applies to all instances that reach the leaf, a set of classifications, or a probability distribution over all possible classifications. To classify an unknown instance, it is routed down the tree according to the values of the attributes tested in successive nodes, and when a leaf is reached the instance is classified according to the class assigned to the leaf (2005, p.62).

One advantage of the decision trees is that they show the represented knowledge in a straightforward way, easy to be followed (depending, of course, on the size of the tree).

The decision tree employed on this work was generated by the J4.8 algorithm, which is a variant of the C4.5 algorithm distributed as part of the Weka package (WITTEN AND FRANK, 2005). On general basis, the C4.5 works in the following way. C4.5 uses

the fact that each attribute of the data can be used to make a decision that splits the data into smaller subsets. C4.5 examines the difference in entropy that results from choosing an attribute for splitting the data. The attribute with the highest difference in entropy is the one used to make the decision. The algorithm then recurs on the smaller sub lists. For entropy, in this context, it is understood how random the distribution of classes in a set of inputs is. Figure 3.4 shows a decision tree for solving the XOR problem.
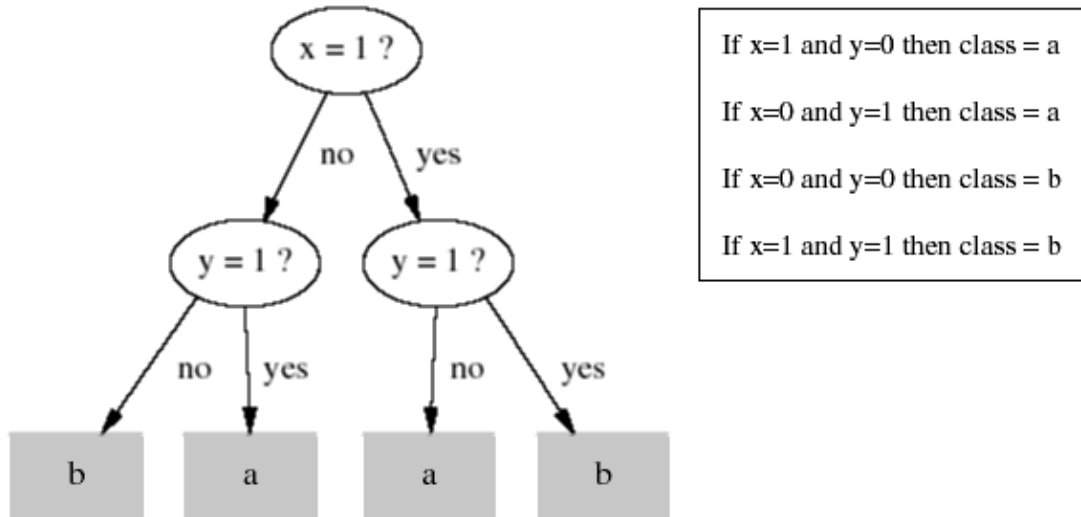
Figure 3.4: Decision tree for the XOR Problem.

3.4 Graph Assembly

After finding the light verb related to a certain syntactic pattern in which a verb is inserted in, a connection between these two verbs is created to indicate their semantic correlation. This forms a graph-like structure, where initially all the verbs are connected with the same weight, indicating the absence of any information about their semantic similarity. These initial weights can be set as non-null values, and a policy of incremental changes in the weights is adopted.

For the purposes of this work, two verbs will be considered "connected" if, between them, there is an edge weighting at least 0.7 and "non-connected" otherwise. The initial weights were set *ad hoc* as 0,5, and a threshold of 0.7 for two verbs to be seen as connected was a choice which tried to balance noise rejection (high threshold values), on one hand, with the reduced amount of inputs on the other hand (low frequency values). A bigger threshold value would lead to less frequent verbs not being considered connected to their respective prototype.

Strengthening (as well as weakening) weight operations were done on two different ways: linearly or exponentially. Both approaches have easy implementation, based upon incremental addition or multiplication, respectively, of reinforcement or punishment terms. Values for the terms must be designated carefully, in order to prevent noise disturbances. Experimental trials have shown that a strengthening factor of 0,085 and a weakening factor of 0,025 provided a good rejection of spurious connections.

For the implementation of the graph, we employed an adjacency matrix. In an adjacency matrix M, two nodes a and b are connected through an edge when M(a, b) = 1. This kind of matrix leaves open other possibilities, such as using the value of M(a,b)

as the weight or strength of the connection between a and b. In this work, we used this strategy. As stated before, all the connections were initialized as 0,5, that is, the entire matrix was filled with this value.

For the establishment of the connections, the following procedure is carried out. For each incoming pair new_verb-*light_verb*, its edge is located and reinforced. The connection of the verb with all other verbs will be weakened. This is carried out for every incoming pair. The resulting matrix is then pruned (using a predefined threshold – for this work, 0.7) and the verbs with connection strength greater than the threshold are considered connected.

3.5 Processing Sequence

For ease of implementation, three different programs were built to perform this work. One program was responsible for extracting the sentences from the database, collecting the features and assembling the feature vectors (the "Feature Detector", regarding the block diagram on figure 4.1). The output of this program was a set of files, containing the feature vectors of each sentence. These files were submitted to the second program, which implemented the decision tree. For each feature vector contained on the file, the decision tree provided the respective light verb (this program correspond to the "Construction Detector", pictured on figure 4.1). The pairs formed by these light verbs and the sentence's verb were further stored in a file, to be processed by the next program, responsible for the construction of the verb graph. This program used each pair verb-light verb to reinforce a connection, on the adjacency matrix, between these two verbs. Each pair (inferred by the previous program) is seen as an evidence of the semantic analogy between the two verbs of the pair. The resulting graph explicitly shows the clusters of semantically related verbs, linked together by means of their appropriate constructions. It can be thought of as the beginnings of the mental lexicon, which is emerging from the data received as input from the linguistic environment to which the learners are inserted. After all pairs are presented to the program, it performs a pruning, in order to cut off weak connections (originated from mistakes on classification). The final output is a list of verbs with their correspondent light verbs, which is used to draw the graph

On the next chapter, we present the experiments performed, as well as the graphs constructed.

4 EXPERIMENTS

In this chapter, a series of experiments is presented, in order to evaluate the ability of the decision tree in identifying syntactic constructions holding some prototypical meaning. The learner found in this step, will, further, be used to allow the construction of the verb graph. Also, with the chosen learner, we will effectively construct the graph using the corpora chosen.

4.1 Learner Experiments

In this section, we carried out a series of tests, centred on a decision tree learner, in order to find its best set of attributes, in order to be used in the further experiments. Our experiments aimed to discover which, and how much, information were needed to accomplish this task.

For these experiments, we chose 3 of the most frequent light verbs, with the prototypical meanings of movement (represented by the verb *go*), caused movement (represented by the verb *put*) and transference (represented by the verb *give*) as our target meanings. The verbs occurring in the sentences were mapped to these verbs through the constructions in which they were inserted.

In each experiment the goal of the learner identify correctly a construction, on the basis of the input data given to it, described in section 3.2.2. All of the input samples were manually labelled with the class that they belong to, that is, we identify to which *light verb* the sentence structure is correspondent. The label GB (acronym for "Go Behaviour") identifies the samples with the syntactic pattern related to the meaning of movement, PB ("Put Behaviour") to caused movement, GVB ("Give Behaviour") to transference and NB ("None Behaviour") samples with none of the previous meanings.

For the purposes of this work, clean data is defined as sentences showing the prototypical meaning of a light verb (go, put, give) along with its respective syntactic pattern. Ambiguous data, on the other hand, are sentences showing the same syntactic patterns of a given construction, but a different meaning.

The reminder of this section will deal with the descriptions of the experiments.

4.1.1 Baseline Experiment

On this section we are going to perform the baseline experiments of our investigations. Our goal here is measuring the amount of constructions that can be

identified on the basis of just a single feature: the type of its complements (NP, PP or VP), in order to obtain a lower bound for further experiments. In this experiment, we are dealing only with clean data, whose distribution is shown in the table 4.1.

Table 4.1: Data distribution for the baseline experiment

Type	Training	Test
GB	500	25
PB	500	25
GVB	500	25
NB	500	25
Total	2000	100

The 500 inputs of each class were, originally, a set of 50 inputs which was repeated 10 times in order to reduce the effect of data sparsity. Using only 50 sentences for constructing the trees, we achieved inconclusive results.

The results for this experiment can be seen in the table 4.2.

Table 4.2 - Results of the baseline experiment.

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	24	1	0	0	25	0,774	0,96	0,857
PB	0	22	3	0	25	0,415	0,88	0,564
GVB	0	24	1	0	25	0,111	0,04	0,077
NB	7	6	5	7	25	1,000	0,28	0,438

In this table (and in the reminder of this work) the precision, recall and f-measure are calculated using the following formulas:

precision = (number of correctly classified inputs / total of inputs classified on that class)

recall = (number of correctly classified inputs / total of inputs of that class)

$$\text{f-measure} = (\beta+1).P.R/(\beta.P)+R$$

As can be seen, the parameter used - type of complement - is a reliable way of positive identification of *go*-constructions. On the other extreme, due to the high degree of variability of this feature, it is useless to identify NB. The distribution of errors is similar to an equally probable distribution. The demand for the same sequence of complements, in many times, for *put*-constructions and *give*-constructions (NP PP) leads to an ambiguity that is not possible to solve only with this information (type of complement), what explains the bad performance in GVB case.

To have a better understanding of how the learner can work with this only feature, we performed a second experiment, divided in steps, in which we used increasing amounts of ambiguity on each step. Figure 4.1 shows the performance of the learner under such conditions. The composition of training and test sets follows the same guidelines described on the section 4.1.3, but in the present case, the feature vectors have only the type of the complements. As the chart shows, even with ambiguity, the GB-class can be reliably identified, probably because is the one with just a single PP as

complement. PB and GVB classes are as expected more difficult to be distinguished due to the ambiguity of its members. The improvement in precision of the NB class is due to the larger number of these inputs in the training set.
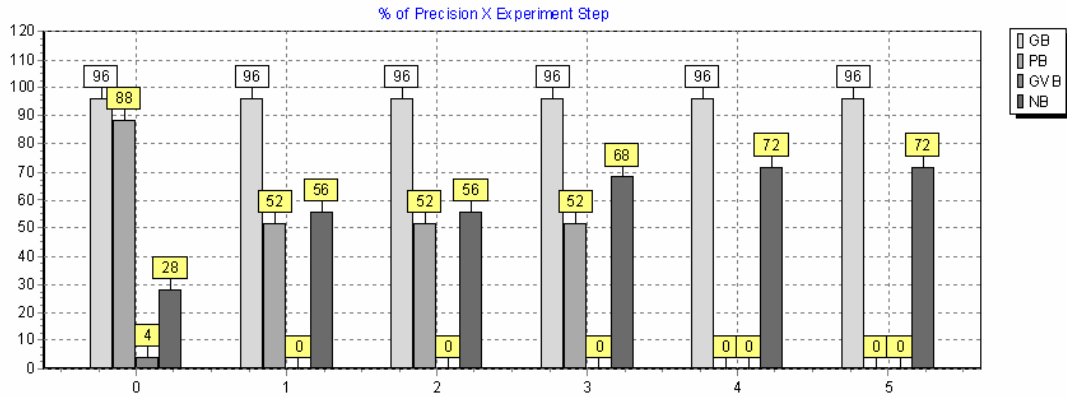
Figure 4.1: Results for the baseline experiment using increasingly ambiguous data

4.1.2 Experiment 1

In this experiment, we evaluate the performance of the decision tree for identifying constructions on the basis of the full range of features collected. We intend to see how better the classification is, in relation to the baseline experiment. Two approaches were adopted for this experiment. In the first approach we replace the value of the feature Head noun POS in the feature vectors (shown in table 3.2) with a more general POS tag, where if the tag is (a) nominal it is changed to N, (b) verbal to V and (c) pronominal to D (*here*, *there* and *that* when they are used as complement e.g. *Put it there*). The second approach uses the original POS-tags. We intend to determine how detailed this information must be to ensure a positive identification. We used the same distribution of training and test inputs as that used in the baseline experiment. This distribution can be seen in the table 4.1. The results for this experiment are shown in the tables 4.3 and 4.4.

Table 4.3: Results for the experiment 1 with general POS-tags

| True Class | Assigned Class | | | | Statistics | | | |
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	24	0	1	1	25	0,857	0,96	0,906
PB	0	23	2	0	25	0,821	0,92	0,868
GVB	0	0	2	23	25	0,286	0,08	0,125
NB	4	5	2	14	25	0,378	0,56	0,452

Table 4.4: Results for the experiment 1 with detailed POS-tags

| True Class | Assigned Classes | | | | Statistics | | | |
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	24	0	0	1	25	0,828	0,96	0,889
PB	0	22	3	0	25	0,786	0,88	0,83
GVB	0	0	14	11	25	0,778	0,56	0,651
NB	5	6	1	13	25	0,52	0,52	0,52

As we can see, the more general tags were not enough to improve the identification of GVB-class inputs, but only made clear that they did not belong to PB-class. So, in this case, the available information is insufficient for a correct classification. On the other hand, it is possible to see the differences caused by the addition of the other information present on the feature vectors. The most salient ones are the improvement of the precision for classification in the classes GVB and NB, as well as a shift of the classification errors of the GVB class, towards the NB class. The additional information enabled the disambiguation between GVB and PB classes.

4.1.3 Experiment 2

Experiment 2 tested the influence of the amount of ambiguous sentences in the input data on the identification of the constructions. To achieve this, we divided this experiment in five stages. For each stage, we had a training set following the same distribution shown on the table 4.1, with is 50 inputs repeated ten times for each class. Also, we had the addition of an increasing amount of ambiguous inputs. At each stage, we added 90 inputs in the class NB (9 inputs repeated ten times). These 90 sentences are divided equally in 30 ambiguous sentences for each of these classes: GB, PB and GVB. In order to have a parameter for comparison, the test set was kept the same, for all trials. The ability of the learners would be reflected on the evolution of the error rates. As the amount of ambiguity increases, we expect degradation on the ability of the learners in correctly classifying a sentence, due to the presence of the same features on different classes, which may lead to mistakes. Table 4.5 details the composition of the training set over five conditions with different ambiguity levels.

Table 4.5: Distribution of the training data for the experiment 2

Composition of the Training Sets					
	Step 1	Step 2	Step 3	Step 4	Step 5
GB	500	500	500	500	500
PB	500	500	500	500	500
GVB	500	500	500	500	500
NB	590	680	770	860	950
% ambiguity	4,35%	8,26%	11,89%	15,25%	18,36%

The amount of ambiguity is calculated as follows:

Amount of ambiguity = (number of ambiguous inputs / total of inputs) x 100%

Again, this experiment was carried out in two modalities, one using general POS-tags and other using detailed tags.
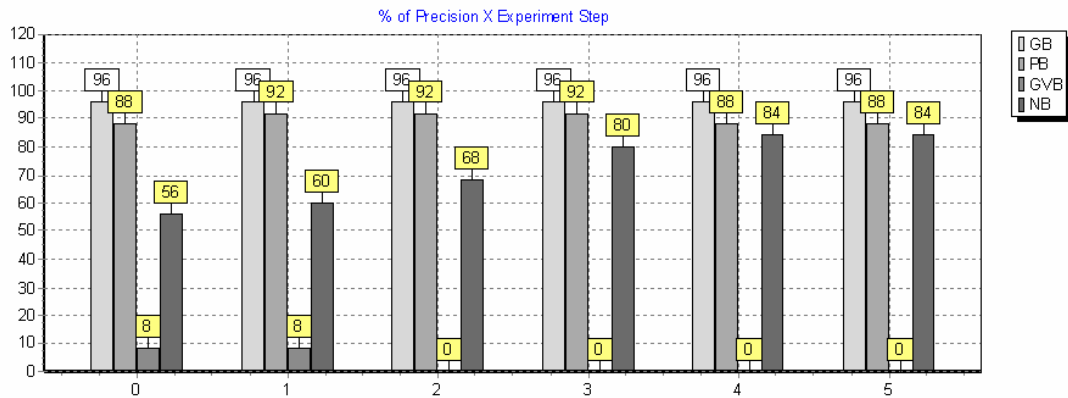
Figure 4.2: Results using general POS-tags
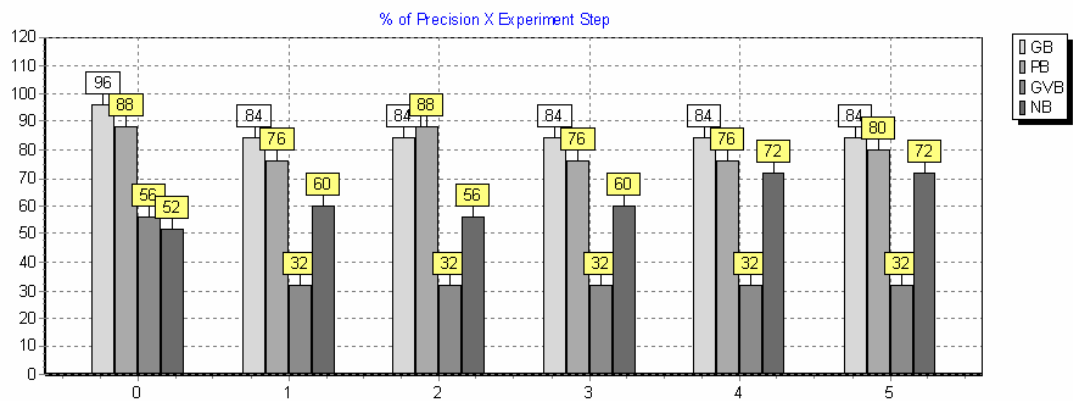
Figure 4.3: Results using detailed POS-tags.

The step 0 for these both experiments represents the experiment 1 using general tags (figure 4.2) and with detailed tags (figure 4.3), respectively, added to the chart for the sake of comparison. The results for the experiment with general tags are shown in figure 4.2. On this chart, we can see the variation of precision from step to step. As can be noticed, the increasing ambiguity poses little changes on the performance rates. Figure 4.3 exhibits the performance of the decision tree with detailed information. Again, small changes can be observed from one step to another. Comparing with the experiment using general tags, we can see a slightly lower precision rate in the classes GB and PB, and a higher precision for the classes and GVB and NB, on average.

4.1.4 Intermediate Considerations

In the last experiments, we could see that the addition of information has improved the ability of the decision tree in identifying the constructions. However, syntactic features did not provide enough information for complete disambiguation. For instance, although the following sentences both have the same syntactic pattern and feature vectors, they denote different events, and just the first one denotes movement:

- Helen walks to her house. (Subj V Obl)

Table 4.6: Example of feature vector with ambiguity

Type of object	Prep.	Head noun POS	Type of object	Prep.	Head noun POS	Type of object	Prep.	Head noun POS	Head categ.
PP	To	NN1	?	?	?	?	?	?	?

- Kelly talks to her mother. (Subj V Obl)

Table 4.7: Example of feature vector with ambiguity

Type of object	Prep.	Head noun POS	Type of object	Prep.	Head noun POS	Type of object	Prep.	Head noun POS	Head categ.
PP	to	NN1	?	?	?	?	?	?	?

Thus, the learners have no means of differentiating between them, and they may incorrectly classify the second in the same class as the first. Looking more closely, we can see that in order to differentiate them, the learner should have further information about the objects of these two sentences, with motion verbs having a place or direction as complement. Therefore, a learner that is based on Goldberg's proposal would require more than only syntactic information to be successful.

Hence, in the next experiments, we investigate if the use of semantic restrictions can improve the disambiguation of syntactically ambiguous instances, and help the learner acquire successfully the mappings between verbs and constructions. In order to do that, we use an additional feature in the feature sets. In addition to the POS-tag of the head nouns, we also added their "types" in the form of their hyperonyms, extracted from Wordnet. Given that any noun in Wordnet is a subtype of more general concepts (with wider degree of abstraction) in the Wordnet hierarchy, gathering information concepts further up in the hierarchy, like "animal", "location" and "person", may be enough to provide the necessary information. With the addition of the semantic information, the feature vectors are as shown in the tables 4.9 and 4.10.

The previous sentences have the following feature vectors:

- Helen walks to her house (PP, to, NN1, "location")

Table 4.8: Example of feature vector enriched with semantic information.

Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.
PP	To	NN1	location	?	?	?	?	?	?	?	?

- Kelly talks to her mother. (PP, to, NN1, "person")

Table 4.9: Example of feature vector enriched with semantic information.

Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.	Type of object	Prep.	Head noun POS	Head categ.
PP	To	NN1	person	?	?	?	?	?	?	?	?

In the next sections, we verify the performance of the learners under the same experimental conditions as before, but now with access to this additional feature containing some degree of semantic information.

4.1.5 Experiment 3

In this experiment, we evaluate the influence of the semantic information in the classification ability of the learner. In the same way as the experiment 1, we followed the same two modalities for this experiment, using the same training and test data from the experiment 1 (whose distribution is shown in the table 4.1), but, this time, enriched with semantic information extracted from Wordnet. The results for the experiment with general POS-tags are shown in the table 4.10 and with detailed POS-tags in the table 4.11.

Table 4.10: Result for the experiment 3 with general POS-tag

	Assigned Class				Statistics		
True Class	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	25	0	0	0	25	0,893	1	0,943
PB	0	23	2	0	25	0,676	0,92	0,78
GVB	0	5	18	2	25	0,783	0,72	0,75
NB	3	6	3	13	25	0,867	0,52	0,65

Table 4.11: Result for the experiment 3 with detailed POS-tags

	Assigned Class				Statistics		
True Class	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	24	0	0	1	25	0,857	0,96	0,906
PB	0	23	2	0	25	0,676	0,92	0,78
GVB	0	6	8	11	25	0,727	0,32	0,444
NB	4	5	1	15	25	0,556	0,60	0,577

As these results show us, they are better than the baseline experiment as expected. We can see, also, several GVB inputs misclassified as PB, which did not happen on the experiment 1. The combination of semantic information with more general POS tags provided, however, the best average result for this experiment, as well as the best rate of positive identifications for the GVB-class, which showed to be quite troublesome. A more detailed analysis of these results will be presented in a further section.

4.1.6 Experiment 4

Like in experiment 2, now we address the effect of the semantic information with ambiguous data on the learner's ability of correctly classifying its input. This experiment followed the same guidelines of the experiment 2, using the same distribution of inputs along the classes, as shown in the table 4.5 and the same test set for all steps as shown in the table 4.1. Again, we performed two modalities of experiments. One of them uses more general POS-tags and another using more detailed tags. Figures 4.4 and 4.5 show the evolution of performance rates for, respectively, the experiment with general POS-tags and detailed POS-tags. Again, step 0 on the charts shows the results of the experiment 3, as a matter of comparison.
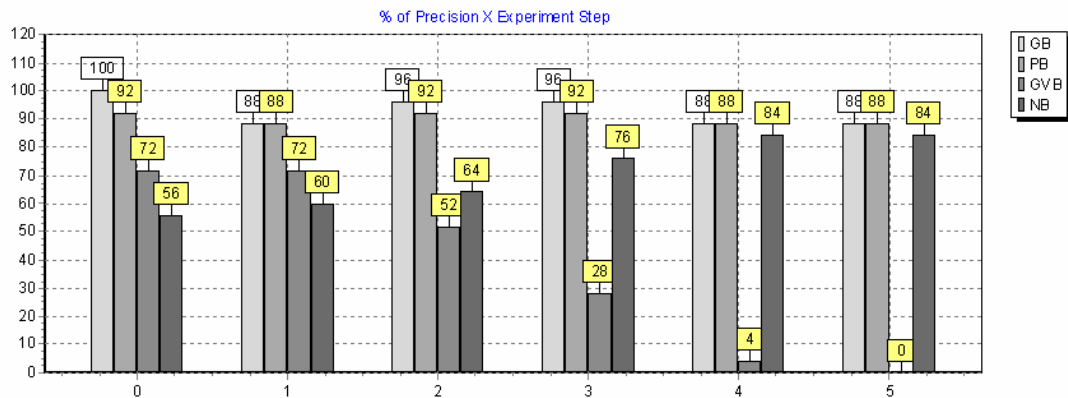
Figure 4.4: Results using general POS-tags and semantic information
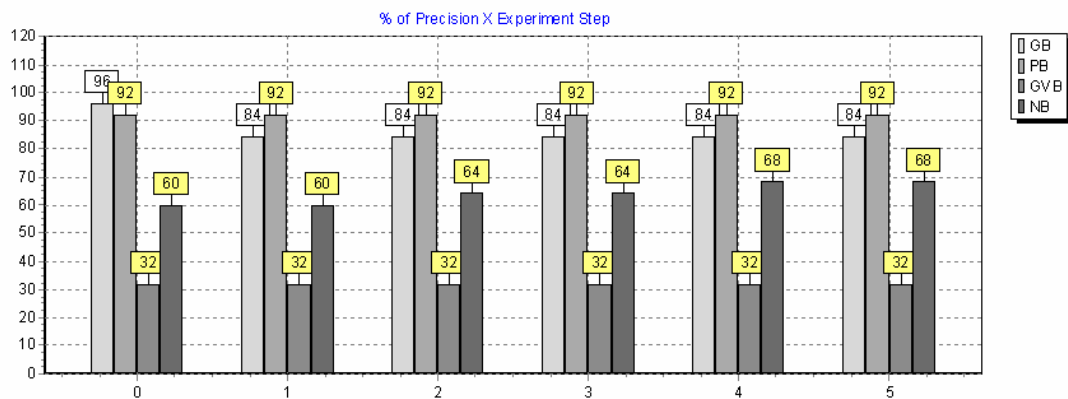
Figure 4.5: Results using detailed POS-tags and semantic information

From the charts, we can see that the influence of ambiguity was negligible in the experiment using detailed tags. On the other hand, with more general tags, the increasing ambiguity has lead to an almost steady increment in error rate, to the point of complete failure for the class GVB. Reminding figure 4.2, we can see the same behaviour. These two experiments have in common the use of general POS tags. We can conclude that these tags do not have enough information to disambiguate and classify correctly the input data. For instance, there was no way to differentiate persons (recipients of give) from objects (all were identified with the tag "N"), which would be necessary to correctly identify transference actions (give), which demand, at least, a person on the construction.

4.2 Discussion

For better evaluating the results obtained, it is important to discuss some points about the input data. As mentioned before, the pre-processing of the input sentences generated tagging and parsing errors. Although not frequent, tagging errors, discussed in section 3.2.2, could not be corrected during the feature extraction step, and some of these errors are reflected in the feature vectors generated. Parsing errors, on the other hand, caused a larger impact. First of all, PP-attachment errors, discussed in the section 3.2.2, caused some incorrectly attached PPs to be mistakenly interpreted as part of constructions. We choose not to correct these inputs (this cleaning could not be done automatically). In spite of it, these data were used in the experiments. In addition, some parsing trees had an irregular format, with some elements such as NPs not being fully identified. These cases returned empty values, which are represented by the symbol "?". As a consequence, not all vectors provide full information about the sentence. Another problem was the identification of head nouns. Although most of them were easily identifiable, others were instances of compound nouns, and demanded some heuristics to be identified. For instance "chocolate cake" and "five dogs" are examples of such troublesome groupings. To decide which noun is the head of the phrase, we defined hierarchy with three levels of preference of POS-tags. The noun with higher level was chosen as the head. In case of two or more nouns having the same preference, the first occurring was chosen. The hierarchy of preference is shown in the appendix B. Lastly, a significant problem arose in the search over the Wordnet for the hyperonym of a head noun due to polysemy. That is, each sense of a head noun had its own hyperonym. As word sense disambiguation was out of the scope of this work we had no manner to define which sense was to be used. Therefore, as many words had more than one sense, we used the first (the most frequent one) returned. This turned out to be a double-edged sword: it inserted information for the system, but inserted noise too. Along with this problem of noisy information, there is another one: missing information. Some features could not be collected from the parsed data (due, again, to irregularities on the parsing trees). Most of the feature vectors have some missing feature, which leads to an incomplete construction of the decision tree and difficulties to the classification of these inputs.

From the whole of the experiments, we could see that there is a limit for the granularity of the information (that is, how much the information is specific) for the construction of the decision tree, as well as the amount of information used. Very few information, as used in the baseline experiment, leads to a very simplistic model, which is not able to differentiate constructions with syntactic similarities, such as *"Tony put the bags (NP) on the floor (PP)"* and *"Sarah sent a letter (NP) to her brother (PP)"*, which are, respectively, *put* and *give* constructions. The use of the all features of the feature vector made possible a higher rate of positive identifications, but the very low frequency of some values and the incidence of noise had their prejudicial effect in the classification. In order to ensure the classification, the attribute values identifying some class must occur with a significant frequency. A low frequency makes possible that small amounts of noise are enough to disturb the classification. The use of more general tags, for instance, increases the frequency of certain values, so, few occurrences of wrong values have small or no impact on the construction of the learner. It is worth noticing the difference in performance between the experiments using generic and detailed tags, and the better results of the former when using semantic information. The further addition of semantic information from Wordnet seemed ineffective in

conjunction with detailed tags. This time, we were facing another problem: excess of information, that is, too much information distributed along a little number of instances.

Taking these considerations into account, it is not surprising that the most successful experiments were those ones that combined the more generic POS-tags with semantic information, providing the disambiguation needed. Analyzing the test set, we found out that the GVB-class sentences that were misclassified had PPY (you) as its first complement head and person as semantic category (using the detailed tags). With the change of all nominal tags for just one, "N", the semantic feature had a higher influence on the decision tree construction and in the classification as well. For instance, the sentence "I'll give you a tranquilizer", has the following feature set (detailed features): ?,?,PPY,person,NP,?,NN1,?,?,?,?,?. With the use of the general tags, the semantic had a greater importance and contributed towards the right classification of inputs like this. From this, we can conclude that there must be a compromise on the depth of information, which cannot be too general to oversimplify the model, neither too detailed, to be affected by data sparseness.

4.3 Graph Construction Experiments

Using the results obtained in the previous experiments, we have built graphs that show the verbs classified within each construction. Our goal is to indicate how the investigation described here could serve as basis for the acquisition of a mental lexicon, where the verbs are grouped regarding to their syntactic and semantic features. The graphs were constructed as described in the section 3.4. We used the results obtained with the decision tree that used semantic information and general POS-tags (experiment 4) for constructing the graphs. The input data was a set of sentences collected from Bates corpus.

As a starting point, all the verbs are connected in the graph. This condition simulates an initial state, in which the learner does not know yet which words are similar among each other. As long as the learner classifies new instances, the links between verbs of the same category are strengthened and the connections between verbs of distinct classes are weakened. To construct the graph, some combinations of reinforcement/weakening factors for the connection between the verbs were considered, as well as the two approaches to perform these connection changes, linear and exponential. The choices were made to keep a compromise between coverage – the inclusion of verbs with low frequency – and noise rejection – minimize the influence of inputs wrongly classified. Keeping a broad coverage with good noise rejection showed to be particularly difficult, because of the low frequency of some verbs, in one hand, and the classification errors, on the other hand. By the end of the graph construction, the connections weaker than a predefined threshold, which are presumed to be due to noise, are automatically pruned from the graph. The low frequencies of some verbs lead us to define, after some tests, the pruning threshold in 0.7. Values greater than 0.7 showed to be excessively restrictive. Under the same principles, we tested many values for the strengthening and weakening factors. The best choice was 0.085 for the strengthening and the 0.025 for weakening. These values make the alteration caused by a spurious pair, originated by a classification mistake, to be overtaken by the – supposed – greater amount of correct pairs. The tests also showed that exponential changes work better then linear changes. For the same pair of strengthening/weakening factors, the use of linear changes lead to a graph with more connections originated from noise. In the case

of polysemy, equal or approximate amounts of occurrences, (since their amount is enough) will lead the verb to be connected with more than one light verb. Figure 4.6 shows the resulting graph.

For the sake of comparison, in figures 4.7, 4.8 and 4.9 we can see the subgraphs generated using Wordnet data, for the same verbs. It is important to notice that the latter were constructed from the static data contained in this electronic resource. In this work, on the other hand, the graph is the result of the dynamic process of automatically acquiring and classifying data.
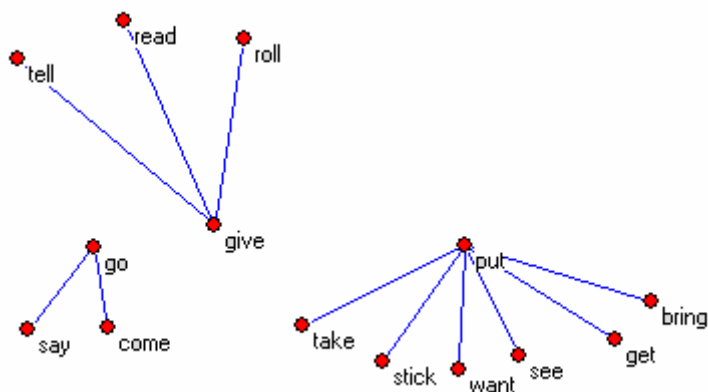
Figure 4.6: Verb graph obtained from sentences of Bates corpus.
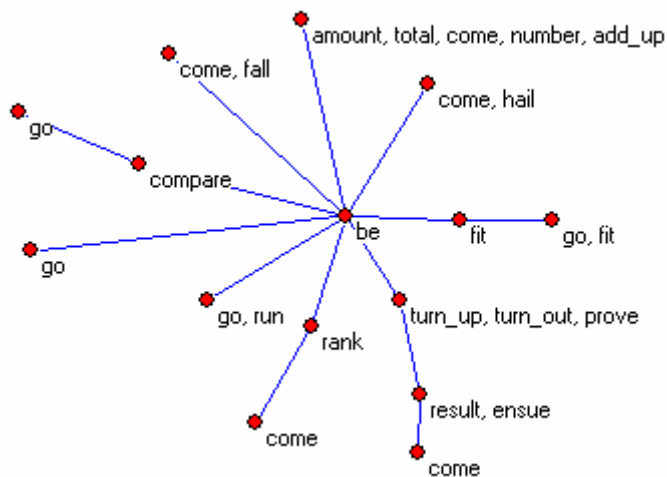
Figure 4.7: Verb graph extracted from Wordnet, showing the connection between the verbs *go* and *come*
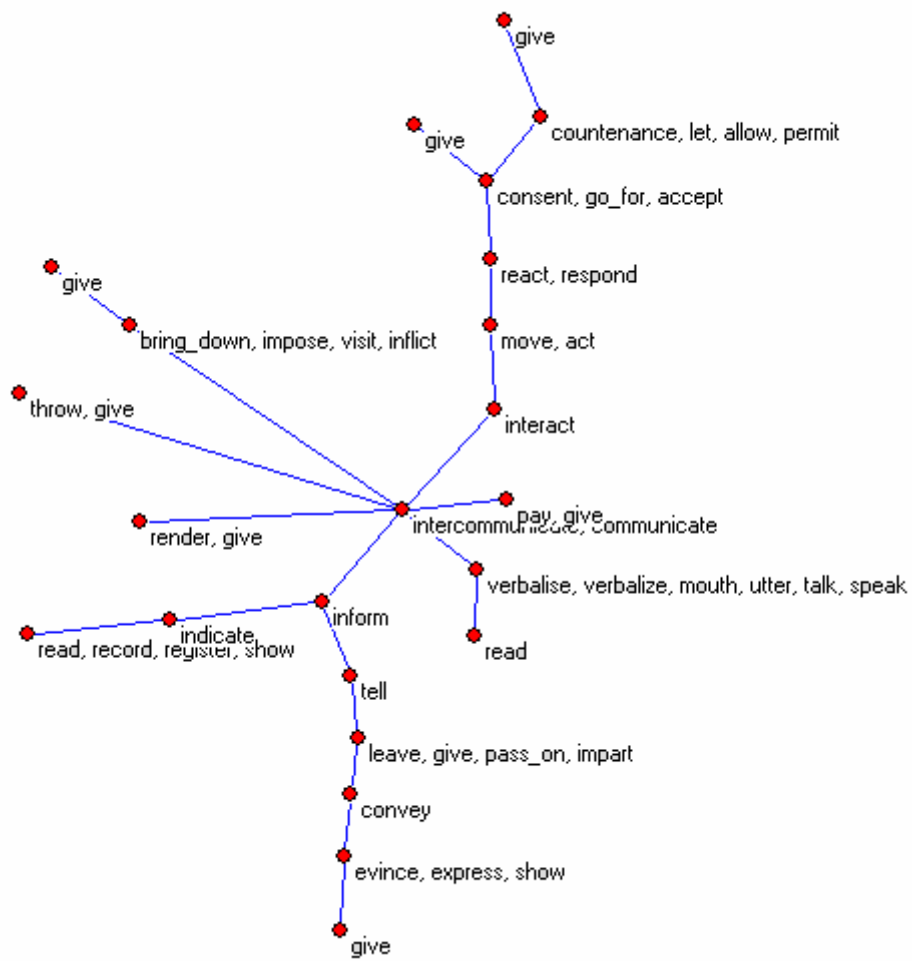
Figure 4.8: Verb graph extracted from Wordnet, showing the connection between the
verbs *give* and *throw, tell* and *read*
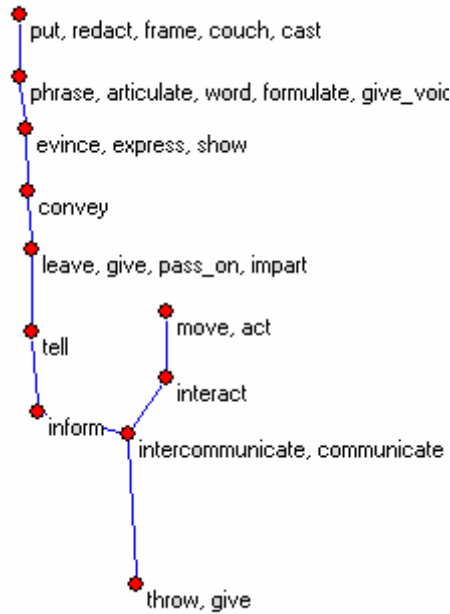
Figure 4.9: Verb graph extracted from Wordnet, showing the connection between the verbs *put* and *throw*.

Comparing our graphs with the ones obtained from Wordnet, we can notice the coherences among them. On one hand, we can see that the connections that we found are, in most part, present in the wordnet, despite the richer details in the later graph. On the other hand, we see a mistakenly connected verb – *want* – in our graph. Although not having any relation with put, their (relatively) similar syntactic patterns and the high number of instances with the verb *want* caused this event.

A basic graph like that resulting from these experiments can be viewed as an initial basis for a more sophisticated and complex (mental) lexical structure, which would emerge as the learning progresses. Therefore, starting from the graph in figure 4.6 the learner, with help of more sources of information, would produce finer-grained distinctions between the verbs in each of the classes, and a hierarchical structure with more intermediate levels would emerge, converging to graphs potentially like those in figures 4.7, 4.8 and 4.9.

5 CONCLUSION

In this work we investigated whether Goldberg's hypothesis - that syntactic constructions have an underlying meaning and that this meaning can be used as a first approximation for the meaning of another verb. This meaning is represented by a particular subset of very frequent and general verbs, the light verbs, which can serve as the basis for determining the construction semantics. The goal was also to determine what would be the requirements for achieving this learning in terms of the environment in which it happens. In particular, we investigated the type of information available to the learner (purely syntactic vs some basic semantic information), and the influence of ambiguous input sentences in the performance of the learners. These results, therefore, can be seen as providing some more evidence not only for Goldberg's proposal, but also for Gleitman and Gillette's of syntactic structures providing cues for the acquisition of the meaning of verbs. We also discussed some limitations and problems with these proposals – such as the need for more information of semantic nature, due to the ambiguity present on the natural language –and investigated whether these could be overcome with the use of some background knowledge about the nature of verbal complements.

We can draw some conclusions from the results achieved along this work. First of all, we could see that was possible to identify *light verbs* – and further connect them with the verbs on the sentences. Apart from performance issues, the results confirmed the claim of syntax as helper in verb learning. Eventual performance shortcomings were not a concern, as long as, playing the role of helper, syntax is not the main, neither the sole, source of information for verb learning. In a completely ideal situation, in which would be possible the integration of (some form of representation of) perceptual information, context, etc the syntactic would not be a problem, as it was in our experiments.

The use of decision trees had its sort of pros and contras. On one hand, its ability to make explicit the acquired knowledge was very attractive, as it made easier to examine the models and explain (to some extent) what was happening. The organization as a tree, with rules clearly bounded, can be followed by anyone with minimal knowledge about the domain under study. On the other hand, decision trees have some difficulties for a work like this. One of them is tied to the advantage previously mentioned: sharply bounded rules just give a "yes" or "no" answer, when a certain level of "perhaps" would be desirable. This feature makes the generalization abilities of decision trees poorer when compared with other models, like neural networks or Bayesian classifiers. Another one is the way in which it is constructed, considering just one attribute at a

time, regardless eventual dependencies or correlations among them. This leads to trees bigger than they should be, and contributes to decreasing of generalization abilities of the trees, as well.

The graph construction technique, on its turn, showed to be effective against moderate levels of noise. However, a strategy which could mimic the cognitive rhythm of acquisition could not be implemented due to time restrictions. It was expected that our graphs would be less detailed than the ones obtained from wordnet. It could not be different, since we were using less information and constructing the graphs automatically.

We must take into account that the results obtained in this work we made very limited use of semantic information. On a real situation, a wide array of information sources is available, such as perceptual inputs, pragmatics and context. All these take place in the acquisition process. Therefore, further more sophisticated distinctions of the verbs acquired can be made as more linguistic and non-linguistic information is incrementally made available to the learner.

These experiments confirm that for learning systems adopting psycholinguistic theories and concepts, such as those of light verbs and constructions, can provide a good basis for language learning, narrowing their search spaces.

5.1 Future Work

From the results obtained, a wide sort of improvements can be proposed. Related to the pre-processing, the use of a different tagger and a different parser could be useful in reducing the amount of error introduced in this step, as well as improving the amount of features collected.

It is interesting, also, the study of other machine learning strategies, to compare with decision trees, such as neural networks.

Another possible improvement is a change of approach for identifying constructions. As presented in this work, we proposed to treat the finding of the initial meaning of a verb as a classification problem. However, it may instead be seen as a clustering process, in which the clusters are created to capture the similarities between the feature vectors describing the sentences, creating clusters as they are needed, unlike a classification task in which the classes are pre-defined. We believe that a comparative investigation of the behaviour of clustering models would allow us to find whether they are more cognitively compatible with human language acquisition.

Moreover, we also plan to use other corpora, with more annotation and more variation on the usage of verbs. Finally, we also want to extend the study to other light verbs, and the evaluation of their syntactic behaviour, firstly looking at how their meanings are acquired in the first place, and then at how these models can perform on the basis of a larger set of possibilities. These results attempt to advance a step further in the implementation of systems capable of on-line learning, which can dynamically acquire and adapt to the language of their environment.

REFERENCES

BATES, E.; BRETHERTON, I.; SNYDER, L. **From First Words to Grammar**: Individual Differences and Dissociable Mechanisms. Cambridge, MA: Cambridge University Press, 1988.

BARRETT, M. Early Lexical Development. In: FLETCHER, P.; MACWHINNEY, B. **A Handbook of Child Language**. Cambridge, MA: Blackwell Publishers. 1995. p. 362-392.

BRISCOE, E.; CARROLL, J. Robust Accurate Statistical Annotation of General Text. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, LREC, 3., 2002, Las Palmas de Gran Canaria. **Proceedings...** Paris: ELRA, 2002. p. 1499-1504.

BUTTERY, P.; KORHONEN, A. Large-scale Analysis of Verb Subcategorization Differences between Child Directed Speech and Adult Speech. In: INTERDISCIPLINARY WORKSHOP ON THE IDENTIFICATION AND REPRESENTATION OF VERB FEATURES AND VERB CLASSES, 2005, Saarbrucken, Germany. **Proceedings…** [S.l.: s.n.]*,* 2005.

CHOI, S.; GOPNIK, A. Early acquisition of verbs in Korean: A cross-linguistic study. **Journal of Child Language**, Cambridge, v. 22, p. 497-530, 1995.

CHOMSKY, N. **Some Concepts and Consequences of the Theory of Government and Binding**. Cambridge, MA: MIT Press, 1982.

DUVIGNAU K.; GAUME B. Linguistic, Psycholinguistic and Computational Approaches to the Lexicon: For Early Verb-Learning. **Cognitive Systems**, Netherlands, v. 6, part 2/3, p. 255-269, 2004.

ELMAN, J. An Alternative View of the Mental Lexicon. **Trends in Cognitive Sciences,** Kidlington, Oxford, v. 8, n. 7, p. 301-306, July 2004.

ELMAN, J. Finding Structure in Time. **Cognitive Science**, Hillsdale, v. 14, n. 2, p.179-211, 1990.

FAURE, D.; NEDELLEC, C. A Corpus-based Conceptual Clustering Method for Verb Frames and Ontology Acquisition. In: WORKSHOP ON ADAPTING LEXICAL AND CORPUS RESOURCES TO SUBLANGUAGES AND APPLICATIONS, LREC, 1998, Granada, Spain. **Proceedings...** Paris: ELRA. 1998. p.5-12. Available at: <citeseer.ist.psu.edu/faure98corpusbased.html>. Visited on: Dec. 2007.

GARSIDE, R. The CLAWS word-tagging system. In: GARSIDE, R.; G. LEECH; G. SAMPSON (Ed.). **The Computational Analysis of English**. Harlow, UK: Lognman Scientific & Technical, 1987.

GENTNER, D. Why Nouns are Learned Before Verbs: Linguistic Relativity versus Natural Partitioning. In: KUCSAJ, S. **Language Development**. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982. v. 2, p. 301-334.

GLEITMAN, L. R.; GILLETE, J. The Role of Syntax in Verb Learning. In: FLETCHER, P., MACWHINNEY, B. **A Handbook of Child Language**. Cambridge, MA: Blackwell Publishers, 1995. p. 413-427.

GOLINKOFF, R. M.; MERVIS, C. B.; HIRSH-PASEK, K. Early object labels: The case for a developmental lexical principles framework. **Journal of Child Language**, Cambridge, v. 21, p. 125-155, 1994.

GOLDBERG, A. K. The Emergence of the Semantics of Argument Structure Constructions. In: MACWHINNEY, B. (Ed.). **Emergence of Language**. Mahwah, NJ: Lawrence Erlbaum Associates, 1999. p. 197-212.

GOLDBERG, A. K. Constructions: a new theoretical approach to language. **Trends in Cognitive Sciences**, Kidlington, Oxford, v. 7, n. 5, p. 219-224, May 2003.

HERBELOT, A. Extracting Entailing Words from Small Corpora for Ontology Building. In: ANNUAL CLUK RESEARCH COLLOQUIUM, 10., 2007, Cambridge,UK. **Proceedings…** Cambridge: [s.n.], 2007.

JACKENDOFF, R. **Semantic Interpretation in Generative Grammar**. Cambridge: MIT Press, 1975.

JOHNSTON, E. **Syllabus for Investigating Minds Lecture**. [S.l.]: Sarah Lawrence College, 1997. Available at: <http://pages.slc.edu/~ebj/IM_97/syllabus.html>. Visited on: Feb. 2007.

LECH, T.; DE SMEDT, K. Ontology extraction for coreference chaining. In: WORKSHOP ON ANAPHORA RESOLUTION, WAR, 1., Newcastle, United Kingdom. **Proceedings...** Tyne: Cambridge Scholars Publishing, 2000. p. 26-38.

LENNEBERG, E. **Biological foundations of language**. New York: John Wiley and Sons, 1967.

MACWHINNEY, B. **The CHILDES project**: Tools for analyzing talk. 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

MANNING, C.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. Cambridge, Massachusetts: The MIT Press, 1999.

MARKMAN, E. **Constraints on Word Learning**: Speculations About Their Nature, Origins, and Domain Specificity. [S.l.]: Stanford University, 2003. Available at: <http://users.ecs.soton.ac.uk/harnad/Papers/Py104/markman.lang.html>.Visited on: July 2007.

MILLER, G. et al. Introduction to Wordnet: An On-line Lexical Database. **International Journal of Lexicography**, Oxford, UK, v. 3, n. 4, p. 235-244, 1990.

PINKER, S. **Language Learnability and Language Development**. Cambridge, MA: Harvard University Press, 1984.

QUINLAN, R. Induction of decision trees. **Machine Learning**, Dordrecht, v. 1, n. 1, p. 81-106, 1986.

REINBERGER, M.; DAELEMANS, W. Unsupervised Text Mining for Ontology Extraction: An Evaluation of Statistical Measures. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, LREC, 4., 2004, Lisbon, Portugal. **Proceedings...** Paris: ELRA, 2004. p. 491-494. Available at: <http://www.cnts.ua.ac.be/Publications/2004/RD04>. Visited on: July 2007

SACHS, J. Talking About the There and Then: the Emergence of Displaced Reference in Parent-Child Discourse. In: NELSON, K. E. (Ed.). **Children's language**. [S.l.]: Gardner Press, 1983. v. 4, p. 1-28.

SKINNER, B. F. **Verbal Behavior**. Englewood Cliffs: Prentice-Hall, 1957.

SMITH, L. B. Learning how to learn words: An associative crane. In: GOLINKOFF, R. M.; HIRSH-PASEK, K. (Ed.). **Becoming a Word Learner**. Oxford: Oxford University Press, 2000.

SISKIND, J.M. A Computational Study of Cross-Situational Techniques for Learning Word-to-Meaning Mappings. **Cognition**, [S.l.], n. 61, p. 39-91, Oct./Nov. 1996. Available at: <http://www.ece.purdue.edu/~qobi/papers/cognition1996.pdf>. Visited on: July 2006

TOMASELLO, M. **Constructing a Language**: A Usage-Based Theory of Language Acquisition. Cambridge, MA: Harvard University Press, 2003.

WIERZBICKA, A. The theory of the mental lexicon. In. HAUSMANN, F. et al. **Handbook of Linguistics and Communication Science**. Berlin: [s.n.], 2004.

WITTEN, I. H.; EIBE, F. **Data Mining**: Practical Machine Learning Tools and Techniques. 2nd ed. San Francisco, CA: Morgan Kaufmann, 2005.

APPENDIX A MORE ABOUT EXPERIMENTS 2 AND 4

In this appendix, we show the tables detailing the results for each step of the experiments 2 and 4.

Experiment 2 with general tags:

Table A.1 - Result for the 1st step of the experiment 2 with general POS-tag

	Assigned Class				Statistics			
True Class	GB	PB	GVB	NB	Total	Precison	Recall	F-Measure
GB	24	0	1	0	25	0,889	0,96	0,923
PB	0	23	2	0	25	0,821	0,92	0,868
GVB	0	0	2	23	25	0,286	0,08	0,125
NB	3	5	2	15	25	0,395	0,60	0,476

Table A.2 - Result for the 2nd step of the experiment 2 with general POS-tag

	Assigned Class				Statistics			
True Class	GB	PB	GVB	NB	Total	Precison	Recall	F-Measure
GB	24	0	1	0	25	0,889	0,96	0,923
PB	0	23	1	1	25	0,891	0,92	0,868
GVB	0	0	0	25	25	0	0	0
NB	3	5	0	17	25	0,395	0,68	0,5

Table A.3 - Result for the 3rd step of the experiment 2 with general POS-tag

	Assigned Class				Statistics			
True Class	GB	PB	GVB	NB	Total	Precison	Recall	F-Measure
GB	24	0	1	0	25	0,889	0,96	0,923
PB	0	23	1	1	25	0,92	0,92	0,92
GVB	0	0	0	25	25	0	0	0
NB	3	2	0	20	25	0,435	0,8	0,563

Table A.4: Result for the 4th step of the experiment 2 with general POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precison	Recall	F-Measure
GB	24	0	1	0	25	0,889	0,96	0,923
PB	0	22	1	2	25	0,957	0,88	0,917
GVB	0	0	0	25	25	0	0	0
NB	3	1	0	21	25	0,438	0,84	0,575

Table A.5: Result for the 5th step of the experiment 2 with general POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precison	Recall	F-Measure
GB	24	0	1	0	25	0,889	0,96	0,923
PB	0	22	1	2	25	0,957	0,88	0,917
GVB	0	0	0	25	25	0	0	0
NB	3	1	0	21	25	0,438	0,84	0,575

Experiment 2 with detailed tags:

Table A.6 Result for the 1st step of the experiment 2 with detailed POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precison	Recall	F-Measure
GB	21	0	0	4	25	0,875	0,84	0,857
PB	0	19	3	3	25	0,731	0,76	0,745
GVB	0	0	8	17	25	0,727	0,32	0,444
NB	3	7	0	15	25	0,385	0,60	0,469

Table A.7: Result for the 2nd step of the experiment 2 with detailed POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precison	Recall	F-Measure
GB	21	0	0	4	25	0,840	0,84	0,840
PB	0	22	2	1	25	0,786	0,88	0,830
GVB	0	0	8	17	25	0,727	0,32	0,444
NB	4	6	1	14	25	0,389	0,56	0,459

Table A.8: Result for the 3rd step of the experiment 2 with detailed POS-tag

True Class	Assigned Class				Statistics'			
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	21	0	0	4	25	0,875	0,84	0,857
PB	0	19	2	4	25	0,760	0,76	0,760
GVB	0	0	8	17	25	0,727	0,32	0,444
NB	3	6	1	15	25	0,375	0,60	0,462

Table A.9: Result for the 4th step of the experiment 2 with detailed POS-tag

True Class	Assign Class				Statistics			
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	21	0	0	4	25	0,865	0,84	0,857
PB	0	19	2	4	25	0,864	0,76	0,809
GVB	0	0	8	17	25	0,727	0,32	0,444
NB	3	3	1	18	25	0,419	0,72	0,529

Table A.10: Result for the 5th step of the experiment 2 with detailed POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	21	0	0	4	25	0.875	0.84	0.857
PB	0	20	2	3	25	0.87	0.8	0.833
GVB	0	0	8	17	25	0.727	0.32	0.444
NB	3	3	1	18	25	0.429	0.72	0.537

Experiment 4 with general tags:

Table A.11: Result for the 1st step of the experiment 4 with general POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	22	0	2	1	25	0.88	0.88	0.88
PB	0	22	2	1	25	0.667	0.88	0.759
GVB	0	5	18	2	25	0.783	0.72	0.75
NB	3	6	1	15	25	0.789	0.6	0.682

Table A.12: Result for the 2nd step of the experiment 4 with general POS-tag

	Assigned Class				Statistics		
True Class	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	24	0	1	0	25	0.857	0.96	0.906
PB	0	23	1	1	25	0.793	0.92	0.852
GVB	1	1	13	10	25	0.813	0.52	0.634
NB	3	5	1	16	25	0.593	0.64	0.615

Table A.13: Result for the 3rd step of the experiment 4 with general POS-tag

	Assigned Class				Statistics		
True Class	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	24	0	1	0	25	0.857	0.96	0.906
PB	0	23	1	1	25	0.885	0.92	0.902
GVB	1	1	7	16	25	0.7	0.28	0.4
NB	3	2	1	19	25	0.528	0.76	0.623

Table A.14: Result for the 4th step of the experiment 4 with general POS-tag

	Assigned Class				Statistics		
True Class	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	22	0	1	2	25	0.88	0.88	0.88
PB	0	22	1	2	25	0.957	0.88	0.917
GVB	0	0	1	24	25	0.333	0.04	0.071
NB	3	1	0	21	25	0.429	0.84	0.568

Table A.15: Result for the 5th step of the experiment 4 with general POS-tag

	Assigned Class				Statistics		
True Class	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	22	0	1	2	25	0.88	0.88	0.88
PB	0	22	1	2	25	0.957	0.88	0.917
GVB	0	0	0	25	25	0	0	0
NB	3	1	0	21	25	0.42	0.84	0.56

Experiment 4 with detailed tags:

Table A.16: Result for the 1st step of the experiment 4 with detailed POS-tag

	Assigned Class				Statistics		
True Class	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	21	0	0	4	25	0.84	0.84	0.84
PB	0	23	2	0	25	0.676	0.92	0.78
GVB	0	6	8	11	25	0.727	0.32	0.444
NB	4	5	1	15	25	0.5	0.6	0.545

Table A.17: Result for the 2nd step of the experiment 4 with detailed POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	21	0	0	4	25	0.84	0.84	0.84
PB	0	23	1	1	25	0.821	0.92	0.868
GVB	0	1	8	16	25	0.8	0.32	0.457
NB	4	4	1	16	25	0.432	0.64	0.516

Table A.18: Result for the 3rd step of the experiment 4 with general POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	21	0	0	4	25	0.84	0.84	0.84
PB	0	23	1	1	25	0.821	0.92	0.868
GVB	0	0	8	17	25	0.889	0.32	0.471
NB	4	5	0	16	25	0.421	0.64	0.508

Table A.19: Result for the 4th step of the experiment 4 with general POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	21	0	0	4	25	0.84	0.84	0.84
PB	0	23	1	1	25	0.885	0.92	0.902
GVB	0	0	8	17	25	0.8	0.32	0.457
NB	4	3	1	17	25	0.436	0.68	0.531

Table A.20: Result for the 5th step of the experiment 4 with general POS-tag

True Class	Assigned Class				Statistics			
	GB	PB	GVB	NB	Total	Precision	Recall	F-Measure
GB	21	0	0	4	25	0.84	0.84	0.84
PB	0	23	1	1	25	0.885	0.92	0.902
GVB	0	0	8	17	25	0.8	0.32	0.457
NB	4	3	1	17	25	0.436	0.68	0.531

APPENDIX B HIERARCHY OF PRIORITY FOR HEAD NOUNS

In this appendix, we present the heuristic used to discover the head noun of a NP. As we had no clue to identify which noun is the head of a given NP, we used the following criteria.

1. If there is just one noun in the phrase, it is the head.

2. If there is no noun, but there is a pronoun, it is considered the head.

3. If there is not any noun nor any pronoun, and, if there is a "here", a "there" or a "that", it is considered the head.

4. If there are two or more nouns, any combination of nouns, pronouns and determiners, that one with more preference is chosen. If there is a tie, the first is chosen. The preference is determined by the POS-tag of the word.

The tables B.1, B.2 and B.3 show the priority regarding to the POS-tag.

Table B.1: Tags with most priority

POS-Tag	Description
ND1	singular noun of direction (north, southeast)
NN	common noun, neutral for number (sheep, cod)
NN1	singular common noun (book, girl)
NN1$	genitive singular common noun (*domini*)
NN2	plural common noun (books, girls)
NNL1	singular locative noun (street, Bay)
NNL2	plural locative noun (islands, roads)
NP	proper noun, neutral for number (Indies, Andes)
NP1	singular proper noun (London, Jane, Frederick)
NP2	plural proper noun (Browns, Reagans, Koreas)

Table B.2: Tags with mild priority

POS-Tag	Description
NC2	plural cited word ('ifs' in 'two ifs and a but')
NNJ	organization noun, neutral for number (department, council, committee)
NNJ1	singular organization noun (Assembly, commonwealth)
NNJ2	plural organization noun (governments, committees)
NNL	locative noun, neutral for number (Is.)
NNO	numeral noun, neutral for number (dozen, thousand)
NNO1	singular numeral noun (no known examples)
NNO2	plural numeral noun (hundreds, thousands)
NNT	temporal noun, neutral for number (no known examples)
NNT1	singular temporal noun (day, week, year)
NNT2	plural temporal noun (days, weeks, years)
NNU	unit of measurement, neutral for number (in., cc.)
NNU1	singular unit of measurement (inch, centimetre)
NNU2	plural unit of measurement (inches, centimetres)
NPD1	singular weekday noun (Sunday)
NPD2	plural weekday noun (Sundays)
NPM1	singular month noun (October)
NPM2	plural month noun (Octobers)

Table B.3: Tags with less priority

POS-Tag	Description
PN	indefinite pronoun, neutral for number ("none")
PN1	singular indefinite pronoun (one, everything, nobody)
PNX1	reflexive indefinite pronoun (oneself)
PPH1	It
PPHO1	him, her
PPHO2	Them
PPHS1	he, she
PPHS2	They
PPIO1	Me
PPIO2	Us
PPIS1	I
PPIS2	We
PPX1	singular reflexive personal pronoun (yourself, itself)
PPX2	plural reflexive personal pronoun (yourselves, ourselves)
PPY	You
DD1	singular determiner (this, that, another)
DD2	plural determiner (these, those)

APPENDIX C RESUMO EXPANDIDO

Um Modelo de Aquisição de Verbos Guiado por Construções Sintáticas

Desde a segunda metade do último século, as teorias cognitivas têm trazido interessantes visões no que tange ao aprendizado de linguagem. O nível de refinamento dos modelos atuais permite explicar fenômenos que, até alguns anos atrás, eram incógnitas. A maior compreensão destes mecanismos, por outro lado, faz possível a criação de modelos computacionais que simulem a aquição de linguagem de maneiras cada vez mais sofisticadas e com melhor performance. A meta deste trabalho é obtenção de um modelo para aquisição de significado de verbos com inspiração cognitiva, Definimos a tarefa de aquisição como o estabelecimento de uma ligação entre um verbo e o seu significado. Este significado é representado por um referente prototípico para a ação representada pelo verbo a ser aprendido.

Aquisição de linguagem não é uma tarefa simples. Embora possa parecer enganosamente fácil, dado que crianças aprendem a usar linguagem desde tenras idades, não se tem ainda uma noção completa dos fenômenos que suportam a aquisição. Reduzindo-se o escopo, e tratando-se apenas da aquisição léxica, e deixando-se de fora a aquisição de estruturas sintáticas abstratas, ainda assim temos uma tarefa de grande vulto. isto porque cada classe gramatical possui suas próprias complexidades no que concerne ao aprendizado. Isso fica patente no caso dos verbos. Diferentemente de substantivos (concretos) os referentes para os verbos (as ações por eles designadas) não possuem limites claros, podem não representar ações observáveis (querer, pensar). Ainda, verbos podem denotar ações em curso, já ocorridas ou por ocorrer. Estes fatores evidenciam a dificuldade da aquisição de verbos meramente com base nas percepções do ambiente. Tais questões foram ressaltadas por Gleitman e Gillette (1995). Elas narram um experimento, no qual detectaram que a identificação de substantivos usando apenas a percepção é muito mais fácil que a identificação de verbos.

Estas premissas deram sustentação para a hipótese de que verbos precisam de alguma forma de "ajuda externa" para serem aprendidos. Para Gleitman e Gillette, bem como outros pesquisadores, esta ajuda externa é provida pela estrutura sintática onde o verbo está inserido, na forma do *syntactic bootstrap*. Nesta hipótese, a estrutura restringe o espaço de busca para o significado do verbo.

Esta conexão entre estrutura sintática também é explorada por Goldberg (1999), que estuda a incorporação de significado em estruturas sintáticas (que formam as assim chamadas *construções*). Ela propõe que, não importando o significado original de um verbo, quando ele ocorre em uma estrutura sintática particular, ele adota o significado inerente a esta estrutura. Por exemplo, o verbo "*buzz*" (zumbir) não tem o significado de

movimento, mas a sentença "*The fly buzzed into the room*" transmite a idéia de movimento. Isto acontece porque a estrutura onde *buzz* está inserido está associada ao significado de movimento. Ela propõe que, inicialmente, o aprendizado de tais estruturas segue uma base verbo-a-verbo, isto é, crianças associam as estruturas de argumentos indivudualmente para cada verbo. Entretanto, devido ao número de verbos em uma linguagem, crianças devem ter um meio de generalizar sobre instâncias particulares previamente aprendidas. Goldberg sugere que o significado construcional está associado a verbos altamente frequentes e gerais, os *light verbs*, como *go*, *put*, *make*, *do* e *give*. Estes verbos são altamente polissêmicos e podem ser usados em uma ampla gama de situações relevantes à experiência humana cotidiana. Goldberg argumenta que, devido à frequência e generalidade, a criança é exposta a esses verbos juntamente com suas respectivas contruções, o que estabelece uma associação entre a estrutura sintática e o significado destes verbos.

Uma série de experimentos foi efetuada para avaliar a viabilidade tanto do *syntactic bootstraping* quanto da teoria das construções como fundamento teórico em modelos computacionais de aquisição de linguagem (no caso, de verbos). Os objetivos eram, basicamente, identificar as construções com base em informações extraídas das sentenças, e mapear os verbos das sentenças para o significado das construções a que estavam associados. Como entrada para os experimentos, empregamos a porção do CHILDES, mais especificamente, dos corpora Bates, Brown e Sachs, os quais são são estudos longitudinais do inglês. Estes corpora contém transcrições de fala direcionada a crianças, com crianças de diversas idades, em diferentes situações e contextos (escola, família, etc). Estes dados foram escolhidos a fim de prover um *setting* mais naturalístico para os learners, lembrando, em certa medida, o ambiente linguístico ao qual uma criança está exposta. As sentenças foram preprocessadas usando um sistema de parsing robusto (Briscoe e Carroll, 2002) com POS-tags obtidas do CLAWS2 tagset (Garside, 1987), como descrito por Buttery e Korhonen (2005).

A partir das informações léxicas e sintáticas disponíveis após o preprocessamento, extraímos características dos complementos de cada verbo nas sentenças. As características usadas foram as seguintes:

- o tipo do complemento (PP, NP ou VP);

- a preposição usada, no caso de uma PP, um marcador de ausência de preposição no caso de uma NP ou VP;

- o núcleo do complemento NP (ou da NP embutida, no caso de PP's);

- a POS-tag do núcleo do complemento;

- categoria semântica do núcleo do complemento, na forma de seu hiperônimo, extraído da Wordnet.

Das características apresentadas acima, apenas o núcleo do complemento não foi usado. O grande número de possíveis substantivos traria uma esparsidade indesejada para este atributo. Entretanto, ele foi usado na obtenção da categoria semântica (hiperônimo do núcleo, obtido através da Wordnet). O conjunto de características extraídas de cada sentença foi organizado como vetores, cada vetor correspondendo aos complementos de um dado verbo.

Após extrair os vetores de características de todas as sentenças, tivemos que etiquetá-los manualmente com uma flag indicando a classe a qual pertencem. Vetores

de características de sentenças de movimento, relacionadas ao verbo go, receberam a etiqueta 'Go-class'. Da mesma forma, para a construção de movimento causado, relacionada a 'put', atribuímos a etiqueta 'put-class' e para a de transferência, relacionada a 'give', usamos a etiqueta 'Give-class'. Para todas as outras construções, fora do escopo deste estudo, usamos a etiqueta 'Others-class'. Para reduzir a insidência de esparsidade nos vetores de características, limitamos em 2 o número de possíveis complementos para cada verbo. Para verbos com apenas um complement, o valor para as características do segundo complemento foram definidas como '?' (valor nulo). Características para as quais não foi possível obter o valor, também atribuímos o valor '?'. Características que não tenham valor (como preposições em NP's, por exemplo) recebem um outro marcador, a fim de fazer claro que a característica não tem valor, em vez de não ter sido possível obtê-lo.

Para o learner, utilizamos uma árvore de decisão (Quinlan, 1996), que modela, como uma estrutura de árvore, um processo de tomada de decisão para a classificação de um conjunto de entradas. Cada ramo representa uma escolha entre um número de alternativas e cada nodo folha representa uma classificação ou decisão. Em particular, usamos o algoritmo J48, (uma variante do algoritmo C4.5), distribuído como parte do ambiente de experimentos Weka (Witten and Frank, 1999). A maior vantagem de usar árvores de decisão nestes experimentos é a forma explícita na qual o conhecimento adquirido é representado e exibido. Outras abordagens, tais como redes neurais e redes bayesianas, podem obter melhor performance em precisão e recall, mas o modo pelo qual elas chegaram a uma decisão não fica claro. O learner foi avaliado usando conjuntos de treinamento e teste montados com distribuições iguais de vetores de cada classe, a fim de se evitar problemas de desbalanceamento. Os vetores foram escolhidos aleatoriamente dentro dos totais disponíveis para cada classe.

A etapa final do trabalho é a construção do grafo, interconectando os verbos das sentenças e os *light verbs* a eles associados. O grafo é representado por uma matriz de adjacências ponderadas, representando a força da ligação entre o verbo e o *light verb* correspondente. O reforço ou enfraquecimento das ligações se dá mediante a identificação da construção onde o verbo está inserido e o reforço da ligação entre o verbo e o *light verb* associado e o enfraquecimento de todas as outras ligações daquele verbo. As taxas de reforço e penalidade froam cuidadosamente escolhidas, de modo a garantir que entradas ruidosas (em frequencia esperada baixa) não interfiram nas ligações legítimas).

Foram realizados 2 modalidades de experimentos: identificação de que atributos seriam adequados para o treinamento da árvore de decisão e experimentos de montagem do grafo de palavras. Na primeira modalidade, foram efetuados 5 lotes de experimentos, a saber:

1. Baseline, com apenas o tipo da phrase no vetor de atributos, cujos resultados serviram como um piso comparativo para os outros;

2. Experimentos sem informação semântica:

   a. Vetores de características com POS-tag da forma como apresentada no CLAWS2 tagset;

      i. Com dados limpos (sem ambiguidade);

      ii. Com quantidade crescente de ambiguidade;

      b. Vetores de características com tags simplificadas, indicando apenas a classe do núcleo do complemento

           i. Com dados limpos (sem ambiguidade);

           ii. Com quantidade crescente de ambiguidade;

3. Experimentos com informação semântica

      a. Idem alínea a do item anterior

      b. Idem alínea b do item anterior

Em linhas gerais, os experimentos usando dados com informação semântica mostraram um resultado melhor do que sem. O uso da POS-tag do núcleo do complemento trouxe vantagens e problemas ao mesmo tempo, pois aumentou a acurácia para algumas classes, mas, devido à sua esparsidade, gerou uma árvore ramificada demais. A melhor combinação encontrada foi o uso de tags reduzidas (informando apenas se o núcleo era um substantivo, verbo, pronome ou determiner).

Definida a abordagem para identificação das construções, todos os vetores de características gerados a partir dos corpora em estudo foram classificados. Disto resultou um conjunto de pares verbo-construção, que foram usados para alimentar o gerador do grafo de palavras. Foram gerados grafos simples, conectando os verbos das sentenças ao light verb correspondente à construção presente na sentença. Na montagem do grafo, inicialmente, todos verbos eram conectados com todos com o mesmo peso. À medida que os pares verbo-*light verb* eram fornecidos, a ligação do verbo com o *light verb* era modificada por um termo de reforço, e todas as outras ligações daquele verbo modificadas por um termo de penalização. Nos experimentos de montagem do grafo, forma testados tanto os valores dos termos de reforço e penalização quanto a estratégia para a modificação (soma e subtração vs. multiplicação de termos). Assim, se chegou a um valor final de 0,085 como fator de reforço e 0,025 como fator de penalização e do uso de multiplicação em vez de soma e subtração.

Esta etapa do trabalho sofreu alguns problemas devido a erros de tagging e parsing nas sentenças, o que levou à dificuldade (ou incapacidade) de obtenção do valor de algumas características, ou ainda, à obtenção de valores errados. O excesso de valores faltantes nos vetores de treinamento também prejudicou o treinamento do learner. Outro problema foi a identificação do núcleo do complemento no caso de substantivos compostos (como *chocolate cake*). Este problema exigiu uma heurística baseada nos POS-tags para ser resolvido. Ainda, a polissemia de algumas palavras trouxe problemas para a obtenção do hiperônimo a partir da Wordnet, dado que muitos eram retornados. Todas estas fontes de ruído devem ser levadas em considração durante as considerações sobre a acurácia do modelo.

Ao final do trabalho, pudemos observar que mesmo apenas com informações obtidas das sentenças, foi possível estabelecer as conexões entre verbos e *light verbs,* bem como observar a relação entre estrutura sintática e significado. Constatamos que as maiores dificuldade ao se lidar somente com informação linguística são a presença de ambiguidades e o desempenho dos recursos de processamento (parsers e taggers), que eventualmente acabam por inserir mais ruído nos dados. Pode-se visualizar vários melhoramentos para este trabalho, tais como:

- O trato de uma gama maior de construções (foram tratadas apenas três);

- A busca por características que melhor identifiquem as construções;
- O uso de outras abordagens de *machine learning* para a identificação das construções,
- O emprego de um corpus mais linguisticamente rico;
- A busca para estratégias para geração de grafos mais ricos em detalhes.

APPENDIX D CLAWS2 TAGSET

In this appendix, we show the entire CLAWS2 tagset, as well as a brief description of each one tag's meaning. Further information can be found in XXX.

TAG	Description
$	Germanic genitive marker - (' or 's)
&FO	formula
&FW	foreign word
(	punctuation tag - left bracket
)	punctuation tag - right bracket
,	punctuation tag - comma
-	punctuation tag - dash
-----	new sentence marker
.	punctuation tag - full-stop
...	punctuation tag - ellipsis
:	punctuation tag - colon
;	punctuation tag - semi-colon
?	punctuation tag - question-mark
APP$	possessive pronoun, pre-nominal (my, your, our etc.)
AT	article (the, no)
AT1	singular article (a, an, every)
BCS	before-conjunction (in order (that), even (if etc.))
BTO	before-infinitive marker (in order, so as (to))
CC	coordinating conjunction (and, or)
CCB	coordinating conjunction (but)
CF	semi-coordinating conjunction (so, then, yet)
CS	subordinating conjunction (if, because, unless)
CSA	'as' as a conjunction
CSN	'than' as a conjunction
CST	'that' as a conjunction
CSW	'whether' as a conjunction
DA	after-determiner (capable of pronominal function) (such, former,same)
DA1	singular after-determiner (little, much)
DA2	plural after-determiner (few, several, many)
DA2R	comparative plural after-determiner (fewer)
DAR	comparative after-determiner (more, less)

DAT	superlative after-determiner (most, least)
DB	before-determiner (capable of pronominal function) (all, half)
DB2	plural before-determiner (capable of pronominal function) (eg. both)
DD	determiner (capable of pronominal function) (any, some)
DD1	singular determiner (this, that, another)
DD2	plural determiner (these, those)
DDQ	wh-determiner (which, what)
DDQ$	wh-determiner, genitive (whose)
DDQV	wh-ever determiner (whichever, whatever)
EX	existential 'there'
ICS	preposition-conjunction (after, before, since, until)
IF	'for' as a preposition
II	preposition
IO	'of' as a preposition
IW	'with'; 'without' as preposition
JA	predicative adjective (tantamount, afraid, asleep)
JB	attributive adjective (main, chief, utter)
JBR	attributive comparative adjective (upper, outer)
JBT	attributive superlative adjective (utmost, uttermost)
JJ	general adjective
JJ	general comparative adjective (older, better, bigger)
JJT	general superlative adjective (oldest, best, biggest)
JK	adjective catenative ('able' in 'be able to'; 'willing' in 'be willing to')
LE	leading co-ordinator ('both' in 'both...and...'; 'either' in 'either... or...')
MC	cardinal number neutral for number (two, three...)
MC$	genitive cardinal number, neutral for number (10's)
MC-MC	hyphenated number 40-50, 1770-1827)
MC1	singular cardinal number (one)
MC2	plural cardinal number (tens, twenties)
MD	ordinal number (first, 2nd, next, last)
MF	fraction, neutral for number (quarters, two-thirds)
NC2	plural cited word ('ifs' in 'two ifs and a but')
ND1	singular noun of direction (north, southeast)
NN	common noun, neutral for number (sheep, cod)
NN1	singular common noun (book, girl)
NN1$	genitive singular common noun (domini)
NN2	plural common noun (books, girls)
NNJ	organization noun, neutral for number (department, council, committee)
NNJ1	singular organization noun (Assembly, commonwealth)
NNJ2	plural organization noun (governments, committees)
NNL	locative noun, neutral for number (Is.)
NNL1	singular locative noun (street, Bay)
NNL2	plural locative noun (islands, roads)
NNO	numeral noun, neutral for number (dozen, thousand)
NNO1	singular numeral noun (no known examples)
NNO2	plural numeral noun (hundreds, thousands)
NNS	noun of style, neutral for number (no known examples)
NNS1	singular noun of style (president, rabbi)

NNS2	plural noun of style (presidents, viscounts)
NNSA1	following noun of style or title, abbreviatory (M.A.)
NNSA2	following plural noun of style or title, abbreviatory
NNSB	preceding noun of style or title, abbr. (Rt. Hon.)
NNSB1	preceding sing. noun of style or title, abbr. (Prof.)
NNSB2	preceding plur. noun of style or title, abbr. (Messrs.)
NNT	temporal noun, neutral for number (no known examples)
NNT1	singular temporal noun (day, week, year)
NNT2	plural temporal noun (days, weeks, years)
NNU	unit of measurement, neutral for number (in., cc.)
NNU1	singular unit of measurement (inch, centimetre)
NNU2	plural unit of measurement (inches, centimetres)
NP	proper noun, neutral for number (Indies, Andes)
NP1	singular proper noun (London, Jane, Frederick)
NP2	plural proper noun (Browns, Reagans, Koreas)
NPD1	singular weekday noun (Sunday)
NPD2	plural weekday noun (Sundays)
NPM1	singular month noun (October)
NPM2	plural month noun (Octobers)
PN	indefinite pronoun, neutral for number ("none")
PN1	singular indefinite pronoun (one, everything, nobody)
PNQO	whom
PNQS	who
PNQV$	whosever
PNQVO	whomever, whomsoever
PNQVS	whoever, whosoever
PNX1	reflexive indefinite pronoun (oneself)
PP$	nominal possessive personal pronoun (mine, yours)
PPH1	it
PPHO1	him, her
PPHO2	them
PPHS1	he, she
PPHS2	they
PPIO1	me
PPIO2	us
PPIS1	I
PPIS2	we
PPX1	singular reflexive personal pronoun (yourself, itself)
PPX2	plural reflexive personal pronoun (yourselves, ourselves)
PPY	you
RA	adverb, after nominal head (else, galore)
REX	adverb introducing appositional constructions (namely, viz, eg.)
RG	degree adverb (very, so, too)
RGA	post-nominal/adverbial/adjectival degree adverb (indeed, enough)
RGQ	wh- degree adverb (how)
RGQV	wh-ever degree adverb (however)
RGR	comparative degree adverb (more, less)
RGT	superlative degree adverb (most, least)

RL	locative adverb (alongside, forward)
RP	prep. adverb; particle (in, up, about)
RPK	prep. adv., catenative ('about' in 'be about to')
RR	general adverb
RRQ	wh- general adverb (where, when, why, how)
RRQV	wh-ever general adverb (wherever, whenever)
RRR	comparative general adverb (better, longer)
RRT	superlative general adverb (best, longest)
RT	nominal adverb of time (now, tommorow)
TO	infinitive marker (to)
UH	interjection (oh, yes, um)
VB0	be
VBDR	were
VBDZ	was
VBG	being
VBM	am
VBN	been
VBR	are
VBZ	is
VD0	do
VDD	did
VDG	doing
VDN	done
VDZ	does
VH0	have
VHD	had (past tense)
VHG	having
VHN	had (past participle)
VHZ	has
VM	modal auxiliary (can, will, would etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (give, work etc.)
VVD	past tense form of lexical verb (gave, worked etc.)
VVG	#NOME? form of lexical verb (giving, working etc.)
VVN	past participle form of lexical verb (given, worked etc.)
VVZ	#NOME? form of lexical verb (gives, works etc.)
VVGK	#NOME? form in a catenative verb ('going' in 'be going to')
VVNK	past part. in a catenative verb ('bound' in 'be bound to')
XX	not, n't
ZZ1	singular letter of the alphabet:'A', 'a', 'B', etc.
ZZ2	plural letter of the alphabet: 'As', b's, etc.