2.1. Introduction

One of the issues that is seldom discussed in courses on Semantics is the methodology used by the different scholars to arrive at their analyzes. Probably, the reason behind this fact is that, until now, most of the methods were introspective, that is, based on the intuitions of the analyst. However, there are potentially many different methodologies that can be followed to extract meaning from expressions, or, alternatively, to try to pin down meaning, to delimit it more precisely so that the different semantic phenomena (e.g., polysemy, synonymy, co-reference, etc.) can be appropriately described and explained. Lately, new methodologies are being applied, thanks to the progress made by Cognitive Science. Cognitive science, the multidisciplinary study of conceptual systems, tries to combine insights from linguistics, anthropology, cognitive psychology, neuroscience, artificial intelligence and philosophy in order to arrive at a richer and more adequate view of cognition (see Figure 2.1). This entails that the methodologies used in each of these disciplines can be applied to the same topics and, quite often, they can offer new views on old topics. Cognitive Science has much to offer to Semantics; in this unit we will review some of these methodologies that are starting to be applied in semantic studies and will no doubt improve the theoretical basis and the rigor of the discipline. This does not mean that everything that was gathered using the classical introspective methods should abandoned or that this type of methodology should be radically banned from semantics. Even in the most experimental of methodologies both types of approaches are normally combined; as it is sometimes said, “Theory without experiment is empty but experiment without theory is blind”. Therefore, we will find many examples of initial explanations based on the intuitions of the analyst which are then subjected to more objective methods in order to confirm or reject the different hypothesis or to choose between different alternative ones.

This unit can be divided into two parts; the first part will examine one of the most classical approaches in the introspective tradition, the decompositional approach of semantic features; the second part will present a number of empirical methods, namely, those based on statistical, psycholinguistic and neurological information.
2.2. Decomposition of meaning: semantic features

A majority of the approaches to the analysis of meaning tend to assume that the meanings of most words are complex, and can be described as formed by different ‘meaning components’. This strategy seems to fit our intuitions, since somehow, we feel that many words do not have unitary meanings, and very often we can identify parts of meanings of words which are the same; that is, we can identify groups of words whose meanings overlap to a certain degree. These ‘overlaps’ are the parts of meaning which all these words share. At the same time, we sometimes can identify how two words are different. For example, man and woman share part of their meaning: we could capture this by saying that they are both [HUMAN]. At the same, they are also different: what makes them different is gender; the first one is [MALE] and the second [FEMALE]. This suggests therefore a agenda of how to proceed: we can analyze the meaning of a word or expression by identifying those components that are shared by several groups, and then by identifying the parts of meaning that distinguish one word from the next one, and perhaps, by saying something about the way in which those components are combined. These different components of meaning have traditionally been called ‘semantic features’.

Semantic feature analysis is tied to the structuralist approach, which is now almost a century old. Structuralists had described phonology using a method of distinctive binary features (namely, mode of articulation and place of articulation) which is still applied today. This system allowed the comparison of the different sounds of a given language among themselves: /p/ and /b/ were distinguished just by one feature:
Unit 2. Analyzing meaning: main approaches

[±voice]. So, /p/ is [+bilabial][+stop][±voiced] while /b/ is [+bilabial][+stop][+voiced]. Other features would be [±dental], [±fricative], etc. This method also allowed the comparison of sounds from different languages, and, in general, turned out to be a very successful method. Thus, it was only natural that the feature approach were adapted to other levels of linguistics, including semantics.

As we have mentioned previously, when applying a feature analysis to semantics, the meaning components are called semantic features (but they have many more names: semes, semantic components, semantic markers, semantic primes). So, for example, girl, woman, sister, wife and queen all share one semantic feature, which could be called [FEMALE], while boy, man, brother, husband and king would share the feature [MALE]. There is also the possibility of having words which are unspecified for this feature, as in child, person, sibling, spouse and monarch.

The features [MALE] and [FEMALE] are also complementary: that means that we do not need them both, since the presence of one of them implies the absence of the other. We could also say that these features are binary. We just need one of them and add a ‘+’ or ‘-’ sign to indicate one or the other. Therefore, [+MALE] is the same thing as [-FEMALE], and [+FEMALE] is the same as [-MALE]. Which of the two features one uses is, to a certain extent, arbitrary. Nonetheless, it is often the case that theories try to decide which is the most basic feature, or use the notion of markedness to decide which one to use. The most normal or basic one is the less marked one, and that is the one chosen.

One of the advantages of feature analysis is that it supplies an easy and transparent method of capturing the meaning structure (similarities and differences) of groups of words by combining the use of several semantic features. The result is called a semantic matrix. This is what we see in Table 2.1

<table>
<thead>
<tr>
<th></th>
<th>[HUMAN]</th>
<th>[ADULT]</th>
<th>[MALE]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Girl</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Boy</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Adult</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Man</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 2.1. Feature matrix for six person terms

Semantic features thus allow us to capture in an economical, compact and highly explicit way semantic relationships such as hyponymy or incompatibility. Semantic features can also be used to formulate selectional restrictions, that is, to specify the semantic constraints of the arguments of a given predicate. Thus, the semantic anomaly
of John drinks rocks can be captured by saying that the verb to drink imposes a selectional restriction on its object: it must be [+LIQUID]. Note that this is not a feature of the verb itself, but of its complement. We can describe many selectional descriptions with the binary feature approach, though there are certain limits, and features often have to be combined with a decompositional analysis of verb meanings, which goes beyond binary features (e.g. Jackendoff 1990, Dowty, 1990).

2.1.1 Problems of binary semantic features

There are quite a few problems with analyzing meaning in this way. The first and most immediate one is that only certain types of words can be fruitfully analyzed using this method. It works with kinship terms, terms referring to male/female/young/adult animals or humans, and a few more. The great majority of word cannot be analyzed in this way.

Then, we have the problem of subjectivity: there is the possibility that the features that one analyst finds are slightly different from features from some other analyst. If both set of features explained the same set of words, we would have no way of deciding which solution is the most appropriate. To a certain extent, this can also be seen in some of the features proposed, like supplying chair with features such as [FOR SITTING], [WITH ARMS], [WITH BACK]. If we admit this, we should admit features like [FOR LYING] for couches, or [FOR PUTTING YOUR FEET ON THEM] for footstools, etc. The bottom line is that there seems to be no restriction in what type of thing a semantic feature should be.

Also, ideally, we should be able to apply this approach to the analysis of a word until we have extracted all its semantic features. We would have an exhaustive and complete feature list of its meaning, and such a list would be the meaning of the word. That is, it would ideally capture all the relevant semantic aspects and characteristics of the word and would be different from the list of any other word. However, this has proved very difficult, if not downright impossible. No matter how many features we list for a given word, there is always a ‘remnant’ of meaning that escapes us. We could try to do this with the word ‘boy’: the list that captures all the knowledge we have about this word would have to be really long, and quite possible, any real attempt would have to be abandoned as not completely satisfactory.

In distinguishing between different types of words, it is quite easy to come up with trivial (and inane) feature analysis: to distinguish between the animal terms bear, pig, rabbit, tiger, donkey, kangaroo we could supply the features [±BEAR], [±PIG], [±RABBIT], [±TIGER], [±DONKEY], [±KANGAROO]. Each animal would have a positive value of its own feature and a negative value on the rest. Sometimes this analysis is disguised by having features which are almost ‘synonyms’ of the word to be analyzed, resulting in
tautological analyses. Thus, horses are [+EQUINE] and dogs are [+CANINE]. This is obviously not an efficient way of analyzing meaning, since it is tautological, and runs into many problems. First, saying that a tiger has the feature [+TIGER] does not really add much to our knowledge of the word, it just repeats its name (the same happens with the horses-equine and dogs-canine examples). Then, saying that a tiger is also [-BEAR], [-PIG], [-RABBIT], [-DONKEY], [-KANGAROO] again does not tell us much about the meaning of the word. Also, the list could be infinite (there are so many things that a tiger is not: a rat, a bird, a fish, a table, a person, an ice-cream, and so on). In general, features should not be *ad-hoc*, that is established to solve a particular, individual problem; and this restricts their use only to groups of words with features that overlap.

Another issue in semantic feature analysis concerns the nature of the semantic features themselves. What is a semantic feature? Are they *primitives* or can they be decomposed into finer distinctions? Which is the level of granularity that will prove adequate for semantic analysis? This is a controversial issue and there are different answers supplied by different scholars.

As a conclusion, binary semantic features can be used to explain certain phenomena, but there are a number of shortcomings that have been identified:

(i) Only a limited range of lexemes can be analyzed (e.g., verbs are difficult to analyze with this approach)

(ii) There are meaning 'residues' which cannot be analyzed (e.g. you cannot capture all aspects of a word using binary features)

(iii) Most kind of meaning relations cannot be analyzed with binary features.

(iv) Binary semantic feature analyses depend to a certain extent on the subjectivity of the analyst

### 2.3. Statistical methods: HAL and LSA

With the advent of computers, methods which were almost impossible some years ago are becoming common practice. For example, there are now methods which use the processing of very large corpora, with millions of words, to extract information from the co-occurrences of the different words. That is to say, these methods construct a representation of the meaning of a word by looking at the *context* in which that particular word occurs. This idea could seem a bit strange at first: trying to find out about the meaning of a word by looking at syntax, at combinations with other words. However, it is roughly based on Wittgenstein’s idea of “meaning as use”. Wittgenstein believed that the meaning of a word or expression was basically dependent on the use we made of that word/expression.

Approaches that attempt to arrive at meaning in this way (extracting it from
“streams” of language) are also called high-dimensional models, for reasons that will become clear presently. Probably the two most popular high-dimensional models are HAL (or Hyperspace Analogue Language) (Lund & Burgess 1996) and LSA (Latent Semantic Analysis) (Landauer & Dumais, 1997). Both these approaches are really computational models of human semantic memory. The way in which they work is the following: they build semantic representations directly from lexical co-occurrences in language, that is from untagged corpora. They do not work exactly in the same way: LSA essentially counts the frequency each word appears in a given context (roughly a paragraph) and then optimizes the dimensionality of this word-by-context matrix (with a mathematical technique called Singular Value Decomposition). HAL moves a 10-word window across a text base and computes all pair wise distances between words within the window at a given instance; the semantic representation of a word is the frequency and distance it appeared before or after every other word in the lexicon. This means that the scope of LSA is a bit larger (sentences and paragraphs) than that of HAL (which is basically words).

We will review HAL with a little more detail to get a grasp of how this proceeds.

2.3.1. An example of a high-dimensional vectorial model: HAL (Hyperspace Analogue to Language).

HAL uses a large corpus of text, and the basic methodology involves tracking lexical co-occurrences within a 10-word moving window that “slides” along the text. Using these co-occurrences, a high-dimensional meaning space of 140,000 lexical dimensions can be developed. Within the moving window, lexical co-occurrence counts are recorded and their value is weighed: it is inversely proportional to the number of words separating a pair of words. Words that are closer together in the moving window get a larger weight. We can see this with an example.

Imagine that your corpus is just this: “The dog barked in the park. The owner of the dog put him on a leash, since he was barking. Ignoring function words (like the, in, etc.), you see that the target word “dog” appears with words like “bark” (two times), park, owner, and leash. If we looked for words that co-occur with cat in other similar corpora, we would probably find that cat co-occurs with a similar set, but not exactly the same: cats have owners, but don’t bark. Look at Table 2.2. and see whether the co-occurrences make sense and whether you can tell by looking at the list of co-occurrences, which words in the left column are more similar to each other.
Figure 2.2 shows a slightly more precise example of the matrix-construction procedure for a five-word window using the example sentence “the horse raced past the barn fell”. In the table (more technically, the matrix), the numbers in rows encode how close did a given word precede another; numbers in columns tell us the distance at which a given word followed another. We can see how this works by looking at Figure 2.2. First, we see what would happen when analyzing “barn”: the immediately previous word (the) gets a 5, the next one gets a 4 (past) and so on. When analyzing the next word, fell, the same is done: the immediately previous word (this time, it is barn) gets a 5, the next one gets a 4 (the) and so on. Note that when we put all this information in a matrix table, we sum the values found. In the row for barn, the word the gets a 6, because it has a 5 (being the word immediately preceding it) and also an extra 1, (because it appears again at the beginning of the sentence, as can be seen in Figure 2.2). The co-occurrence numbers for this sentence appear in Table 2.3.
In our small example, with a 5-word window, the vector for the word *barn* would be the five digits of the row plus the 5 digits of the column, that is <0 2 4 3 6 0 5 0 0 0 0>. In the real model, they use a 10-word window, and the matrix is actually 70,000 squares. Therefore, in HAL vectors for words have 140,000 dimensions.

Once we have a vectorial representation for each word, there are a number of things that can be done. For example, we can say how close in meaning two words are. By comparing the two vectors, we can get a numerical measure of the semantic closeness of two words. There have been tests in which the HAL system performed the same task given to human subjects. For example, computing the similarity of two given words, or classifying a group of words into different categories.

### 2.4.2. Pros and cons of high-dimensional models

The adequacy of these models as a reflection of human knowledge has been established in different ways. For example, their scores overlap with those of humans on standard vocabulary and subject matter tests; they mimic human word sorting and category judgments, and simulate word-word lexical priming data. Word and discourse meaning representations derived by LSA have been found capable of simulating a variety of human cognitive phenomena, ranging from the acquisition of vocabulary to sentence-word semantic priming and to judgments of essay quality.

The fact that the measures produced by these methods correlate well with several human cognitive phenomena involving association or semantic similarity must be the result of the way peoples' representation of meaning is reflected in the word choice of writers, and/or vice-versa, that peoples' representations of meaning reflect the statistics of what they have read and heard.

Now, these models induce their representations of the meaning of words from analysis of text alone. None of their knowledge comes directly from perceptual information about the physical world, from instinct, or from experience with bodily functions and feelings. Thus the representation of reality we get is bound to be
somewhat sterile and bloodless.

However, it does take in descriptions and verbal outcomes of all these processes, and so far as people have put such things into words, or that their words have reflected such matters unintentionally, these methods have at least potential access to knowledge about them. That is why the authors of LSA in one paper have considered that the knowledge of the world these models have could be seen as analogous to a well-read nun’s knowledge of sex (a level of knowledge, they say, often deemed a sufficient basis for advising the young).

2.4. Psycholinguistic methods

Psycholinguistic methods look at people’s cognitive processing when understanding language (or some other language-related task). These methods usually take some experimental measure to extract conclusions about the mental representation constructed by a given word or expression. These measures can be the reaction time (RT) taken by some task, lexical decision tasks or memory measures. For instance, we can use reading times to examine the difficulty associated with understanding texts; or we can use memory measures to examine what people recall after reading a text, and therefore understand how these texts were interpreted. In this section we are going to examine some of these experimental measures. Specifically, we will look at lexical decision tasks, memory measures and reading times.

2.4.1. Lexical decision and naming measures

In a lexical decision task, participants are presented with a certain letter string on the screen, and their task is to decide (e.g., by pressing a key) whether that letter string makes up a word or not. In a naming task, participants have to read out loud a word presented on the screen. The materials in lexical decision and naming studies typically consist of 50% real words, and 50% non-words (e.g., flurp). The rationale behind both tasks is that people should be faster to identify or pronounce a word when the concept denoted by that word has been made accessible in previous context. Response latencies are therefore thought to reflect the mental accessibility of a word. For example, in a seminal study, Meyer and Schvaneveldt (1971) demonstrated that people were faster to decide that butter is a word in English when it was preceded by the word bread than when it was preceded by the word nurse. As we have seen, this kind of facilitation is normally called priming, and it is attributed to automatic spread of activation between related meanings in the mental lexicon. In the context of lexical decision and naming studies, the first term is often referred to as the prime, and the second term is referred to as the target. Facilitation is said to occur when response times to a target term are
faster after an experimental prime than after a control prime. Inhibition (or suppression) is said to occur when response times to a target term are slower after the experimental prime than after the control prime.

In experiments, there are two parameters that can be manipulated; one is the type of prior context; the other one is the interval that is given between the termination of the priming stimulus and the presentation of the target stimulus. This is often called the inter-stimulus interval or ISI (another name is SOA, for Stimulus-Onset-Asynchrony). Manipulations of ISI are interesting when we want to find about the on-line construction of meanings, the time-line that meaning construction follows. A variation of this measure is the Stimulus-Onset-Asynchrony, or SOA, which is the time which separates the beginning (or onset) of the prime stimulus and the beginning of the target stimulus.

![Figure 2.3. Structure of a lexical decision priming task](image_url)

One example of a study that used this methodology is Blasko and Connine (1993). They used a lexical decision task to examine the time course of metaphor comprehension. Specifically, they were interested in finding out how quickly people
interpret metaphors, and whether the initial stages of metaphor comprehension involve the construction of a literal-related meaning, as implied by the standard pragmatic model of metaphor comprehension (e.g., Searle, 1979). For example, they presented participants with a statement that contained the phrase “hard work is a ladder”, and then participants made a lexical decision either to a literally-related target (rungs), a metaphorically-related target (advance) or a control target (pastry). They found that after participants heard familiar metaphors, response latencies to both metaphorically-related and literally-related targets were facilitated. This finding was interpreted as showing that the comprehension of familiar metaphors does not require the construction of the literal meaning of the utterance before the metaphoric meaning is derived. However, for unfamiliar metaphors the results were different; here only literally-related targets were facilitated in the initial stages of comprehension.

The usefulness of varying the ISI in a lexical decision task can be seen in a study by Till, Mross, and Kintsch (1988). They investigated the time course of sense creation in discourse context. They presented participants with paragraphs that contained statements such as The servant lit the fire and then prepared the meal. He seasoned it and wrapped it in foil. The presentation of the paragraph was then immediately interrupted, and participants were given a lexical decision to a critical target word. In one condition, the target was lexically associated with a word that appeared in the sentence (e.g., tin), and in another condition, the target word reflected an inference from the statement (e.g., barbecue). By manipulating the time between the termination of the sentence and the presentation of the lexical decision target, the authors found that lexically-associated targets were facilitated fairly early after reading the statements (~ 200 msec), but inferentially-related targets were facilitated only later on; about 500 msec after reading the statements, establishing the possible time that should elapse before inferences are drawn.

2.5.2. Memory measures

The logic that underlies the study of language by looking at memory performance is that people’s memory for a certain expression could indicate of how that expression was mentally represented at the time that it was committed to memory, or encoded. In particular, the errors that people make in a memory task may be indicative of the representations they constructed. The basic format of memory tasks consists of two stages: a learning stage in which certain expressions are presented to participants, and a test stage which evaluates participants’ memory for those expressions. Between the learning and the test stage there is typically a break of a certain period in which participants are engaged in an unrelated task to clear working memory. The evaluation of memory during the test stage can be carried out by asking people to recall the items
presented in the learning stage (free recall), or by presenting them with old and new test items and asking them to indicate for each item whether it had appeared in the learning stage (this is known as an old/new recognition task).

The recognition measure is particularly useful when the goal is to examine the degree to which different expressions are similar in meaning. The greater the similarity in meaning between two expressions, the more likely people are to mistake one for the other in a recognition test. For instance, a certain hypothesis might be that statement $a$ is more similar in meaning to statement $b$ than to statement $c$. To test this hypothesis, statement $a$ might be presented in the learning stage of a study, and either statements $b$ or $c$ would be presented in the test stage. Ideally, participants should judge items $b$ and $c$ as “new” items, i.e., as items that had not appeared in the learning stage. However, if the results show that item $b$ is incorrectly judged as "old" reliably more often than item $c$, then the results would support the hypothesis that statement $b$ is closer in meaning to $a$ than statement $c$.

One example of this methodology is Bransford, Barclay, and Franks (1972). In that study, participants studied sentences, and were later given a recognition task in which they were to say whether a certain sentence was presented in the learning stage. The critical materials in the study were ones in which a sentence presented in the learning stage and a sentence presented in the test stage either reflected the same situation in the world, or not. Take for examples, sentences (1) and (2), which describe the same situation in the world:

1. *Three turtles rested on a floating log and a fish swam beneath them.*
2. *Three turtles rested on a floating log and a fish swam beneath it.*

The authors found that when participants were presented with sentence (1) in the learning stage, they later tended to confuse it with sentence (2) in the test stage; i.e., they incorrectly judged that sentence (2) was presented in the learning stage. In contrast, when the two sentences did not describe the same situation in the world, participants did not tend to confuse them. For example, participants who were presented with sentence (3) in the learning stage seldom mistook it for sentence (4):

3. *Three turtles rested beside a floating log and a fish swam beneath them.*
4. *Three turtles rested beside a floating log and a fish swam beneath it.*

These results demonstrated that people are more likely to confuse two statements when the statements refer to the same situation in the world, and indicated that the mental representation of a statement is not just a propositional representation of the phrase structure of the text.

Another example is the study by Boroditsky, Ham & Ramscar (2002) which we
presented in Unit 2, and which tested Indonesian and English speakers. The idea in that study was that the memory of an event in a specific tense would be influenced by the type of language spoken by the subject; Indonesian speakers, having no tense in their language, would focus less on this aspect and would therefore have a worse recall of the correct tense of the event they had seen previously, as was the case.

2.5.3. Reading Times

In some experiments, the time needed to read a text is taken to be indicative of the processing difficulty that is associated with the comprehension of that text. Reading times can be particularly useful for examining whether preceding contexts have differential effects on text comprehension. In such designs, a given text appears in different experimental conditions that vary in the type of context that precedes the text.

An example of a study which uses this methodology is Gentner, Bowdle, Wolff, and Boronat (2001) which examined whether the comprehension of metaphorical statements activates the source domain of the metaphor. They constructed two sorts of experimental conditions. Both conditions ended with the exact same statement, e.g., a statement drawn from the mapping A DEBATE IS A RACE. In one condition (the consistent condition), the preceding statements in the paragraph were instances of the same cross-domain mapping. In the other conditions (the inconsistent condition), previous statements were instances of a different mapping, e.g., A DEBATE IS A WAR. The authors found that the last statement in the paragraph was read faster in the consistent condition than in the inconsistent condition. These findings suggest that when people read metaphorical statements, they activate conceptual structure that corresponds to the mappings between the base and target domains, and the accessibility of this knowledge assists the comprehension of later statements if they are based on the same mapping.

These are some of the main psycholinguistic methods; there are others, such as item recognition tests (you read a test and then are asked about an item of it) or self-report measures (e.g., listing features; in a feature-listing task, participants are typically asked to write down properties that best capture the meaning of a certain expression).

Very often it is useful to separate these methods into two different classes. A group of them, such as lexical decisions, naming, and reading times, rely on measuring behavior that occurs when participants are actively engaged with the experimental material. They are often referred to as “online measures” and are thought to measure processes that are relatively low level and automatic rather than strategic ones. Other methods, such as those that rely on memory, feature listing or fragment completion are “offline measures” as they rely on information that is provided well after the comprehension of the materials has been completed. These measures are therefore more susceptible to the working of strategic considerations employed by participants in
the study, a fact that should be taken into account by careful researchers.

2.5.4. Eye-tracking methods

More recently, psycholinguistics has incorporated more sophisticated equipment in the attempt to measure how people process different cognitive events. One of these methods is **eye-tracking**: a machine that will follow the gaze of subjects as they are performing some tasks and record different parameters of their looking behavior. The perceptual field of our eyes covers 200°; yet, only 2° are surveyed with sharpness and precision, and the rest of the visual field is not attended. This small part of our visual field is the one examined by the fovea, a tiny part of the retina. To cover the rest of the visual field we have to move our eyes quickly from one part to another. These quick movements we make with our eyes are called “**saccades**”; our eyes usually stay about 200-300 ms on a stimulus. Saccadic movements occur three or four times per second. The parameters that psychologists measure in saccades are: the pattern of locations the eye follows (**scanpath**), the duration existing between a stimulus and the saccadic movement (**saccade latency**) and the amount of time spent looking at a particular location (**fixation duration**).
Eye movements are useful because they can give us clues of on-going processes (not just of results) and are ecologically valid (they don’t interrupt the task). They are also a very sensitive measurement method.

Of course, eye-tracking is a methodology that can be used to study many aspects of perception and of cognition and action, not only language. It has been applied to many activities (specially with the use of new equipments, head-mounted eye-tracking, which allow subjects to walk and carry out tasks in the world in natural settings).

Regarding language, the most apparent, and most studied, link between eye movements and language is in the process of reading. The general characteristics of eye movements during reading have been studied in great depth over the past quarter century (for thorough reviews, see Rayner, 1978, 1998). This methodology has revealed a number of important facts about how people’s eyes move when they read. For example, the eyes rest in fixation for approximately 200-250 milliseconds during reading. Saccades between fixations cover an average about 2 degrees of visual angle, although this is better expressed here in terms of a span of 7 to 9 letter spaces, since the number of letters covered remains largely invariant despite differences in text size or distance (Morrison & Rayner, 1981). The chances of an individual word being fixated vary according to whether it is a content word (85%) or a function word (35%) (Carpenter & Just, 1983), and in relationship to the length of the word, with 2-3 letter words being skipped 75% of the time, but 8 letter words fixated almost always (Rayner & McConkie, 1976).

Although readers typically move their eyes forward when reading, approximately 10-15% of saccades move backward, fixating previous letters or words. These regressive saccades are thought to be related to difficulties in processing an individual word, or difficulties in processing the meaning or structure of a sentence; in these cases, readers can often accurately re-fixate the part of the text that generated confusion (Murray & Kennedy, 1988). Details of the cognitive processes of pronoun resolution
and co-reference, word frequency, lexical ambiguity, syntactic ambiguity, as well as the influence of semantic and discourse context on these processes, can all be extracted from analyses of eye-movement patterns (Rayner, 1998; Tanenhaus & Trueswell, 1995). One example of a study using eye-tracking methodology is Spivey & Tanenhaus (1998). They had their subject sitting in front of a display of objects such as a candle, a bag of candy, a pencil, and a spoon. The participants were then instructed to “Pick up the candy”. About a third of the time participants fixated the candle for a couple hundred milliseconds before looking to and reaching for the candy. Participants typically denied looking to the candle at all, and yet their eye movements revealed a process substantially different from their conscious report and their manual action. This kind of brief interference between similar sounding object names occurs not just for cohorts (words that have the same beginning) but also for rhymes (Allopenna, Magnuson, & Tanenhaus, 1998).

Another study used eye-tracking to prove that subjects really construct a mental image when understanding language (instead of translating to some sort of amodal Mentalese). Spivey and Geng recorded participants’ eye movements while they were listening to sentences and were facing a large white projection screen that took up most of their visual field. One of the sentences was: “Imagine that you are standing across the street from a 40-story apartment building. At the bottom there is a doorman in blue. On the 10th floor, a woman is hanging her laundry out the window. On the 29th floor, two kids are sitting on the fire escape smoking cigarettes. On the very top floor, two people are screaming.” While listening to the italicized portion of this passage, participants made reliably more upward eye-movements; this was later on replicated with “downward stories” as well as leftward and rightward stories. In each case, the saccades of subjects coincided with the descriptions supplied in the stories. Thus, while looking at ostensibly nothing, listeners’ eyes were doing something similar to what they would have done if the scene being described were actually right there in front of them.

More recently, Altmann and Kamide (2004) have demonstrated that participants will look at a picture of a cake before hearing the word spoken in the sentence ‘The boy will eat the cake’. This anticipatory saccade will not occur when subjects hear ‘The boy will move the cake’. This evidence demonstrates that participants are activating rich ‘thematic role’ knowledge (Ferretti, McRae, & Hatherell, 2001) of the verb “eat”, and fixating likely candidates for this action before the word is spoken.

2.5 Methods from neuroscience

2.5.1 Event-Related Potentials (ERPs)

Finally, there is another method for gathering information about mental processing that is at the same time more technologically advanced and at the same time more “low-
level” that the previous ones. The study of Event-Related Potentials, or **ERPs**, tries to link language use to the patterns of neuronal firing in our brains. As it is known, when neurons communicate, the signals that travel through the networks are both chemical (with the interchange of different chemical substances from neuron to neuron in synaptic connections) and electric: a synapse between two neurons generates a change in the electric potential of these cells. These synaptic currents can be monitored by placing at least two electrodes somewhere on the head and amplifying the voltage difference between them. The resulting EEG (electroencephalogram) observed at the scalp is due to the summed (postsynaptic) potentials of multiple neurons acting in concert. The changes in potentials that are related in time to a sensory, motor or cognitive event are known as **event-related potentials** (ERPs). A single ERP is too small to be of much help; what is done is to repeat a given event several times and to average the responses that one gets. The result is one voltage waveform in time at each recording site, consisting of negative and positive going oscillations (relative to the pre-event voltage activity). Different studies place a different number of electrodes throughout the scalp, from 20 to 60, to achieve a topographic map of the activity of the brain. One of the great advantages of this method lies in its ability to capture phenomena with a great speed resolution, i.e., very quickly.

![Figure 2.5. Recording Event-Related Potentials](image)

ERPs are especially useful for the study of language comprehension because a negative component peaking around 400 ms after stimulus-onset (known as the N400)
has been shown to vary systematically with the processing of semantic information. This electrophysiological response was first observed by Kutas and Hillyard in response to a semantically anomalous word in a sentence context, such as *city* in the sentence ‘He shaved off his mustache and city’. To date, it has not been observed to incongruities in other settings, such as music, nor to anomalies in language that are non-semantic in nature such as grammatical violations or language-irrelevant changes in the physical attributes of words. Syntactic or grammatical violations affect another different peak, the P600 (a positive wave that occurs 600 milliseconds after the stimulus has been presented).

The amplitude of the N400 to a particular word is highly sensitive to the immediate context in which it occurs, whether that context is a single word, a sentence, or a discourse. For example, N400 amplitude varies with semantic relationships between individual words in lists. Prior occurrence of an associatively related item (e.g. *bee*) or a semantically related item (e.g. *sugar*) reduces the N400 amplitude to a given word (e.g. *honey*) yielding a ‘semantic priming’ effect.

This N400 reduction is similar in time course and scalp distribution to that observed to words in sentences, where N400 amplitude has an inverse relationship to the word’s rated ‘cloze probability’ – that is, the proportion of individuals who provide that particular word as the most likely completion for that sentence fragment in a paper and pencil test. N400 effects thus seem to reflect constraints arising from several levels of context; words that are easier to process because they are expected in a context or are related, semantically, to recently presented words elicit smaller amplitude N400s relative to the same words out of context or in weak or incongruent contexts, respectively.

ERPs can also provide information about issues of representation, for example, about the nature of meaning representation in the brain. One primary question is how information associated with different types of stimuli (abstract and concrete words, pictures, faces, sounds) are stored in and accessed from memory. The two options are
(i) there is a single, common ‘amodal’ store accessed equally by all meaningful stimuli or
(ii) each type of stimulus access its own, modality-specific knowledge base. In this
sense, some studies have shown that concrete words (that is words referring to
pictureable objects), elicit N400s with a more anterior distribution, much like that seen
to line drawings and photographs, than those elicited by abstract words, at least in the
absence of strong contextual information. This and N400 data from other studies
provide a picture which is consistent with neuropsychological data and results from
other neuroimaging techniques in suggesting that semantic memory may consist of
clusters of features which are distributed across multiple, higher-order perceptual and
motor processing areas: the ‘meow’ of a cat, for example, is represented in auditory
association cortex, the shape and color of a cat in the visual cortex, the ‘furiness’ of a
cat in higher-order somatosensory areas, and so forth. The N400 data go further still,
however, showing that meaning emerges from these distributed systems by virtue of
temporally coincident and functionally similar activity within a number of brain areas.
The semantic representation of a unified concept, then, would involve a distinct pattern
of activation, in the N400 time window, across multiple brain areas, reflecting the
contribution of each – for example, more visual for concrete words and pictures, more
auditory for environmental sounds – plus activation in multimodal processing areas that
would serve to ‘bind’ this distributed information together.

2.5.2 Hemodynamic methods (MRIs & PETs)

The electrochemical signals that underlie ERPs create cellular energy demands that are
met by increased blood flow to active areas and increased levels of oxygenated blood in
the area of active tissue. Imaging methods such as positron emission tomography
(PET) and functional magnetic resonance imaging (fMRI) can localize such
metabolic/hemodynamic changes with a much higher level of spatial resolution than
can be achieved by electromagnetic measures.

Although these hemodynamic-based methods are thus useful for determining
which brain areas change their activity levels during experimental manipulations, none
of them have the temporal resolution (on the order of tens or hundreds of
milliseconds) needed to track the use of semantic memory during language
comprehension in real time. The hemodynamic response not only lags behind the
eliciting electrical activity by 1–2 s but also is temporally extended across several (5–15)
seconds even for a brief electrical impulse and multiple events occurring within
milliseconds or seconds will add to each other and give a non-usable signal. Even with
the development of methods for undoing this sort of overlap (such as event-related
fMRI), hemodynamic methods are by their nature limited in their temporal resolution
to events occurring at no less than a second. They remain, however, very efficient for
finding out *where* something is happening in the brain.

As can be seen in Figure 2.7, there have been studies trying to test the reality of Lakoff and Johnson’s theory of metaphorical representation. Rohrer (2004) has measured fMRI responses to sentences containing the word ‘hand’ in a literal sense (e.g., *I handed him a beer*) or a metaphorical sense (e.g., *I handed him the project*). His results show a great degree of overlap in the brain areas activated when processing both sentences, which this author interprets as proof that the embodiment theory of meaning (i.e., that abstract metaphorical thought has a sensorimotor base) is correct.
2.6. What lies ahead: future work

As we have seen, there are many methods that are complementing the traditional linguistic methods and can provide an empirical base to semantic theories. The era of computers has brought along a myriad of methods which were before available only to few laboratories. Today, anyone with a computer can set up experiments in which reading times and priming are studied, or complicated processing of corpora. Other methods, such as eye-tracking or ERPs, remain more expensive and therefore more restricted, though their cost has also come down. Yet other methods, such as neural network modeling, have proved useful, but the extent of their usefulness must still be tested more thoroughly, since there are disagreements on the validity of the conclusions we extract from such a methodology.

The request to back up theoretical explanations empirically also forces theories in the direction of an increased level of explicitness; only a theory that is explicit enough and makes predictions that can be falsified can be submitted to experimental/empirical verification. Some theorists complain that this is too stringent a requirement; there are theories which are accepted as scientific, and yet, do not make predictions or are difficult to falsify. Perhaps the most popular example would be
Darwin’s theory of natural selection and the origin of species. As one biologist said As the evolutionary scholar Waddington observed:

...Natural selection, which was at first considered as though it were a hypothesis that was in need of experimental or observational confirmation, turns out on closer inspection to be a tautology, a statement of an inevitable although previously unrecognized relation. It states that the fittest individuals in a population (defined as those which leave most offspring) will leave the most offspring (Waddington 1960: 385-386)

The movement we seem to be witnessing in these moments is for a mixture of methods. The notion of converging evidence seems to be the key notion here. As a matter of fact, we could say that Cognitive Science as a discipline depends on this notion; the idea seems to be that using different types of methodologies we can advance step by step in the incredibly complex topic of human cognition and meaning.
2.6. Bibliography

The main references used in this unit have been:


