1 Introduction

In contrast to the theorizing and model building in linguistic semantics, psycholinguistic research has not yet led to unified theories of how language users process words and larger chunks semantically. Rather, psycholinguists have concentrated on specific semantic phenomena, for example how ambiguous words are interpreted on-line, and have tended not to use their findings to support general theories of semantic processing. In addition, the cross-pollination between linguists and psycholinguists is still, regrettably, rather limited, with psycholinguists rarely relating their findings to contemporary theoretical semantics, and theoretical linguists paying little attention to relevant experimental results.

In this chapter, we discuss a range of psycholinguistic findings that should be interesting to linguists and could inform linguistic theories. These are, obviously, not the only topics that psycholinguists have studied, and readers can find much more information in, for example, the Handbook of Psycholinguistics (Traxler and Gernsbacher 2006, e.g., chapters 10, 11, 14, 15, 18, 20, 21, 23), the Oxford Handbook of Psycholinguistics (Gaskell 2007), and the Cambridge Handbook of Psycholinguistics (Spivey et al. 2012). The bulk of the present chapter will be devoted to studies of semantic comprehension, but we also make some reference to research on language production, a field that has been growing steadily over the last 20 years or so. One issue that needs addressing is to what extent language comprehension and production should be integrated. Most models keep production and comprehension quite separate, but sometimes assume that they share representations but involve inverse mappings between these representations.

2 Incrementality

One of the most important features of language processing is that it is incremental (e.g. Just and Carpenter (1980); Marslen-Wilson (1973)), with comprehenders starting to interpret each word as soon as it becomes available. Hence, language users do not wait until, for example, the end of a clause (see Fodor et al. (1974)) before they start interpreting the text. This can be seen, for example, in the processing of lexical ambiguities where it is commonly found that an immediate selection is made between the different meanings of a word like bank (Rayner and Duffy (1986); see detailed discussion below). For the initial interpretation of a word, language users can only use contextual (and possibly situational) information
that has already been processed, together with the knowledge that has been stored about that word. In addition, more extensive semantic interpretation is extremely fast, given that people normally understand what they are reading at speeds of about 300 words per minute (Rayner et al. 2012). This suggests that the semantic system must be flexible to quickly deal with a large variety of input. However, this does not necessarily mean that all words are immediately interpreted fully. A classic counterexample comes from the Moses illusion (Erickson and Mattson 1981), in which participants are asked questions such as *How many animals of each kind did Moses take on the Ark?* A staggering 81% of the respondent incorrectly answered “two.” Such mistakes are not due to participants not knowing the story. It was also not the case that the respondents failed to process the incorrect word or did not pay attention to it: in the original experiment by Erickson and Mattson, respondents read the questions aloud themselves; and Reder and Kusbit (1991), using self-paced reading, showed that respondents spent slightly more time processing the incorrect word than the correct version before answering incorrectly. If readers had immediately activated all relevant information about Moses, or about what it means to be a survivor in *Where should we bury the survivors?* (Barton and Sanford 1993), then such errors should not have occurred.

The same research also suggests that the depth of processing can be affected by linguistic factors. For example, when *Moses* is the topic of the sentence, error rates drop to 41% (Erickson and Matteson 1981), Experiment 2; see also Sturt et al. (2004) and Ward and Sturt (2007) for evidence of discourse focus on depth of processing). This indicates that the amount of lexical information that is activated, or at least used immediately, can vary, so that processing can be relatively shallow or more detailed, depending on syntactic structure and possibly the task at hand (see also Ferreira and Patson 2007). This kind of flexibility in the extent that language users activate relevant information is not normally part of linguistic models, though any model that purports to be psychologically valid will need to be compatible with the idea of flexible incremental processing. Of course, this does not mean that linguistic theories have always ignored contextual effects. For example, in Relevance Theory (Sperber and Wilson 1995; Wilson and Sperber 2004; see also Carston 2013), it is assumed that comprehenders make use of contextual clues in order to determine the writer’s or speaker’s intent (see Chapters 10 and 11).

There is other evidence for partial, rather than full, incremental interpretation. For example, while one meaning of a noun-noun ambiguity is selected immediately (see below), readers seem to slightly delay this process in the case of noun-verb (*a/to duck*) and verb-verb (*pen letters/sheep*) ambiguities (Boland 1997; Frazier and Rayner 1987; Pickering and Frisson 2001; see also Folk and Morris 2003). Second, quantifier resolution can also be delayed. In a study using event-related (brain) potentials, Urbach and Kutas (2010) first established that a sentence such as *Farmers grow worms* elicited a greater N400 amplitude compared to a sentence such as *Farmers grow crops*. This effect is associated with semantic anomaly and indicates an immediate problem with the interpretation of *worms*. They then constructed sentences with a quantifier so that the acceptability of the two sentences switched: *Few farmers grow worms* was more acceptable than *Few farmers grow crops*, and *Farmers rarely grow worms* was more acceptable than *Farmers rarely grow crops*. Hence, one would expect that if readers immediately used the contextual information to its fullest extent during on-line processing, the *Few farmers grow worms* sentence would not result in a processing cost (i.e. no N400 effect) but the *Few farmers grow crops* would. When they tested these modified sentences, the N400 effect was attenuated, indicating an early influence of context, but, crucially, the effect did not reverse: *worms* still elicited a larger N400 amplitude than *crops*. This result suggests that the fully specified contextual meaning of the sentence fragment,
as exemplified by the acceptability results, had not yet been computed (see also Filik et al. (2005) for eye-tracking evidence). Third, the finding that readers spent longer processing the last word of a sentence (a so-called wrap-up effect) may reflect additional, higher-order integrative processing of the sentence as a whole (Just and Carpenter 1980). This suggests that interpretation of words in sentences, and how sentences relate to each other, is not always accomplished at the earliest possible moment (see Frazier (1999) for a similar argument).

3 Predictability

While it is clear that language users very quickly integrate incoming information into an unfolding interpretation (incremental processing), there is also good evidence suggesting that people don’t just wait for information to come in but actively predict what might come next.

During reading, readers move their eyes to bring text into the fovea (2 degrees in the centre of vision, approximately 6–8 letters), as it has the highest acuity. However, although the acuity is less, they can also extract some information from the parafovea (about 5 degrees to the left and right from the centre of vision, approximately 15–20 letters on each side). It’s a matter of (strong) debate in eye movement research as to exactly what type of information (letter shape, orthographic, phonologic, morphological, and/or semantic) can be extracted from outside the foveal region (see Rayner (2009) for an overview).

Eye movement research has shown that words that are predictable from the preceding context are fixated for a shorter time and/or are more likely to be skipped altogether (e.g. Balota et al. (1985); Drieghe et al. (2005); Ehrlich and Rayner (1981); Frisson et al. (2005); Gordon et al. (2012); Rayner and Well (1996)). This indicates that readers must have either predicted the upcoming word or processed it parafoveally (i.e. without looking directly at it) before fixating or skipping that word, and that when the target word corresponds to the predicted word, processing is facilitated. Exactly why predicted words are often skipped remains a matter of debate, with some researchers arguing for an “educated guess” strategy in which only coarse information about the upcoming word is processed (e.g. Brysbaert and Vitu (1998)) and others claiming that the upcoming word can be fully processed when the reader fixates towards the end of the preceding word (e.g. Reichle et al. (2003)).

While evidence suggests that words that often co-occur are processed faster than words that do not co-occur (Arnon and Snider 2010; McDonald and Shillcock 2003a, b), predictability effects are due to more than mere co-occurrence frequencies or lexical priming between words close in a sentence. For example, Fitzsimmons and Drieghe (2013) showed comparable predictability effects when the degree of predictability was determined by information directly preceding the target word (hairy preceding spider in 1) or information coming from a preceding sentence (2):

(1) Bill has always been a fearful person. He screamed when he saw the hairy spider in the bath.

(2) Bill is scared of eight-legged creatures. He screamed when he saw the spider in the bath.

Evidence from ERP experiments also indicates that language users can quickly predict upcoming words. DeLong et al. (2005) presented sentences one word at a time while measuring electrical brain activity. They used sentences such as:

(3) The day was breezy so the boy went outside to fly a kite/an airplane.
The noun *kite* is highly predictable, as determined by a cloze task (in which participants completed the sentence fragment up to the critical noun phrase with the first word(s) that came to their mind), while *airplane* is plausible but much less predictable. Previous work had shown that words that were less predictable elicited a larger N400 amplitude than words that were more predictable (Kutas and Hillyard 1984). What DeLong et al. (2005) found was that this N400 effect was already noticeable when participants read the indefinite determiner (*a* or *an*), before they had seen the noun. Thus, the presence of the determiner *an*, which does not fit with the phonology of the predicted word *kite*, resulted in a significantly larger N400 amplitude than was the case when the determiner *a* was presented (see also van Berkum et al. (2005)).

More evidence for the pre-activation of specific upcoming information comes from the so-called Visual World paradigm in which a (typically static) scene is presented on a computer screen (e.g. a scene with a boy sitting in a room, with a ball, a cake, a toy car, and a toy train set around him; cf. Altmann and Kamide (1999)) and a sentence is presented auditorily. Eye movements are recorded during the sentence presentation. Altmann and Kamide (1999) found that participants, when hearing the sentence *The boy will eat the cake*, already started fixating the target object *cake* by the offset of the verb and, thus, before the noun itself was presented. This indicates that the participants had used the verb’s selectional restrictions (i.e. that one eats something edible) to winnow down the possible candidates in the visual scene.

Further experiments have shown that it is not (only) the lexical association between the verb *eat* and the noun *cake* that drives these anticipatory eye movements to the target. Kamide et al. (2003) found that world knowledge information related to the subject was quickly integrated as well. When a scene was presented with two possible agents (e.g. a girl and an adult man) and two possible candidates for the object slot (e.g. a beer and a candy), participants started fixating the most likely object upon hearing the verb. Concretely, more looks to the beer occurred when the sentence started with *The man will taste . . .*, whereas more looks to the candy occurred for *The girl will taste . . .*. In addition, real-world contingencies, e.g. that the use of a past tense might indicate that the nature of an object has changed (e.g. that a pile of feathers is a good representation for birds when hearing *The cat has killed all the . . .*; cf. Altmann and Kamide (2007); see also Knoeferle and Crocker (2007)), are integrated very quickly as well.

The evidence suggests that people employ a variety of sources to predict upcoming information quickly and process that information readily if it indeed appears. Importantly, this upcoming information is often semantic (though people also predict other aspects of language such as phonology). However, many questions remain (see Altmann and Mirković (2009) for discussion). For instance, do people predict continuously (i.e. in advance of every single word of a sentence)? Do people use all sources of information to make predictions and if so, which information takes precedence (e.g. visual vs. lexical vs. auditory input)? And what happens if an unpredicted word appears (is it more costly to process?) and does it matter how closely the word is related to the predicted word? In addition, we shall see below that evidence about the processing of words with multiple meanings suggests that there are limits to the influence of context on the selection of word meanings.

4 Lexical semantics

For a long time, psycholinguistic research in lexical semantics was largely restricted to the processing of semantically ambiguous or homonymous words – words that share orthography (or phonology) but have distinct meanings. (In the interest of clarity, we use homonymy to refer to a single word-form with separate, semantically non-overlapping meanings and
polysemy to refer to a single word-form with semantically overlapping meanings, which we will call “senses”. Some papers have, confusingly, used polysemy to include homonomy. (See Chapter 13.) The central research question, which is related to the modularity debate (see Fodor (1983)), was whether comprehenders selectively accessed contextually appropriate meanings or whether they initially accessed all meanings and then used context to select among them (see Swinney (1979)).

Most evidence suggests that both bottom-up (the activation of all the word’s meanings) and top-down (contextual influence) processes interact during the recognition of homonyms. Good evidence that bottom-up processes are involved comes from the observation that the frequency of the different meanings affects comprehension. For example, in (4) coach is a homonym, with the two meanings (bus and trainer) having roughly equal frequency (Rayner and Duffy 1986); in (5), the word cabin is unambiguous.

(4) He found the coach too hot to sleep in.
(5) He found the cabin too hot to sleep in.

Readers spent longer fixating coach than cabin before encountering the rest of the sentence. It therefore appears that they activated both meanings of coach and competition ensued.

When a neutral context precedes a biased homonym, which has one meaning that is of substantially higher frequency than the other, the dominant meaning becomes available first and no competition ensues. For example, port has a frequent meaning referring to a harbour and an infrequent meaning referring to wine, whereas soup is unambiguous (Rayner and Duffy 1986).

(6) Last night the port had a strange flavor.
(7) Last night the soup had a strange flavor.

Readers spent the same amount of time fixating port and soup before processing the rest of the sentence. This suggests that the two meanings of port did not compete. However, when the following information related to the less frequent meaning of port, as in had a strange flavor interpretation difficulties ensued, which suggests that readers had initially interpreted port as expressing the more frequent harbour meaning.

However, a preceding context affects ambiguity resolution in a different way (Duffy et al. 1988). When preceding context supports the disfavoured meaning of a biased ambiguous word such as port, immediate difficulty ensues. This suggests that the context provides support for the less frequent meaning so that this meaning becomes available at more or less the same time as the more frequent meaning (which is activated in a bottom-up manner). Since both meanings are now available, competition arises. In contrast, when the preceding context supports one meaning of a balanced homonym such as coach, this meaning receives a boost, making it available before the other meaning and the difficulty observed in neutral contexts disappears.

The so-called subordinate bias effect observed for biased homonyms, with extra processing when the subordinate meaning is intended, is very difficult to overcome, even in strongly biasing contexts (e.g. Kambe et al. (2001); Sheridan et al. (2009)), though selective access has been observed in special contexts (e.g. Wiley and Rayner (2000)). These findings about homonym resolution are compatible with the Reordered Access Model (Duffy et al. 1988). Such results pose a challenge to models that assume interpretation depends on lexical co-occurrence, for example as operationalized in Latent Semantic Analysis (LSA; Landauer and Dumais 1997), which have trouble dealing with the representation of multiple meanings.
(McRae and Jones (2013); but see Griffiths et al. (2007)) and for models that posit that meaning is in effect an averaging over episodic traces a word has appeared in (e.g. Kwantes (2005)). In LSA, a word’s meaning is thought of as a matrix containing information about all the contexts it appears (and does not appear) in. In practice, this workspace or dimensional representation is then reduced to calculate the similarity between texts a word appears in. This allows a calculation of the similarity of two words which do not often appear together in the same text (e.g. synonyms). However, given that matrices are constructed for specific word forms, homonyms are collapsed to a single point and the two meanings cannot be distinguished from each other.

While homonyms show clear and immediate effects of both context and frequency, the number of homonyms is in fact quite limited, at least in English and most related languages. In contrast, the vast majority of words are polysemous; for example, book can refer to the object or the content, school can refer to an edifice or an institution. Are these words processed in the same way as homonyms? Most evidence suggests that they are not.

First, the results of lexical decision experiments, in which participants have to respond as quickly as possible whether a letter string presented on a computer screen constitutes a real word or not, indicate that polysemous words are processed differently from homonyms. While older research suggested that ambiguous words are responded to faster than (polysemous and non-polysemous) control words (e.g. Hino and Lupker (1996); see Rodd et al. (2002) for a critical evaluation), newer research specifically contrasting ambiguous to polysemous words showed a processing advantage for the latter (Rodd et al. 2002; Klepousniotou and Baum 2007; see Rodd et al. (2004) for a simulation in an attractor network). Second, data from ERP experiments show different effects for the two types of words, with N400 effects reflecting an effect of dominance/frequency for homonyms, which is absent for polysemous words (Klepousniotou et al. 2012). This means that about 400 ms after word onset electrical brain response signals to homonyms show an influence of the frequency of the different meanings, but no such frequency modification is found for the different senses of a polyseme. Third, studies using magnetoencephalography (MEG) offer neurophysiological support for such meaning/sense distinctions (Beretta et al. 2005; see also Pylkkänen et al. (2006)). And fourth, eye movement experiments have shown distinct processing patterns for homonymous and polysemous words (see below).

In contrast, evidence from more off-line tasks, such as categorization of phrases and sen-sicality judgements, suggests that individual word senses might be processed and represented in the same way as the individual meanings of a homonym (Klein and Murphy 2001; 2002). For example, Klein and Murphy (2001) asked participants to indicate whether phrases such as liberal paper and yellow lecture made sense or not. The previous (prime) trial could contain the same polysemous noun, either in the same sense (daily paper) or a different sense (wrapping paper). They found faster reaction times when prime and target involved the same sense compared to when they did not, and this pattern was comparable to that found for homonyms when the same meaning was re-used (commercial bank followed by savings bank). However, Klein and Murphy did not test for relatedness of the senses and quite a few of their “polysemous” items might in fact be homonyms (i.e. have unrelated meanings) rather than polysemes (i.e. have related meanings). For example, they labelled nail (gun and polish), navy (blue and recruit), and coat (winter and paint) as polysemous even though the relatedness of the two interpretations is far from certain. In addition, it is unclear whether this task reflects early lexical activation or later integration processes (for a more in-depth discussion, see Frisson (2009)).

Several eye movement experiments have examined the processing of polysemous words during normal reading, though only two studies compared polysemy to homonymy directly.
Frazier and Rayner (1990) used words such as book in either their object or content sense and found that, in contrast to the pattern for homonyms, subsequent disambiguation towards the less frequent sense was no more difficult than disambiguation towards the dominant sense. Pickering and Frisson (2001), using ambiguous verbs, found earlier frequency effects for homonyms than polysemes as well as longer processing times for homonyms. In addition, a number of studies looked at the processing of different types of polysemous words by themselves. In a series of experiments (Frisson and Pickering 1999; 2007; McElree et al. 2006), we found that metonymic senses such as the institution sense of school, the event sense of Vietnam, or the product sense of Dickens, were processed as fast as their literal counterparts, at least when the metonymic senses were known by the reader. In addition, the frequency of the individual senses did not affect immediate processing (though later processing measures, arguably reflecting more in-depth processing, sometimes did show an influence of frequency). Recently, Foraker and Murphy (2012; Experiment 3) confirmed the lack of an early processing difficulty for polysemous words such as cotton (the fabric or the crop).

Results from studies about homonymy and polysemy are instructive as they relate to important distinctions in lexical processing and representation. Frazier and Rayner (1990) interpreted their results as reflecting a difference in semantic commitment when processing homonymous and polysemous words, with comprehenders committing immediately to a specific interpretation only when the alternative interpretations are semantically unrelated, as is the case for homonyms. Frisson and Pickering’s (1999) results suggest that for polysemous words, a semantically underspecified meaning is initially activated, which can then be fleshed out or become more specified in later processing (see Frisson and Pickering (2001) for discussion). According to this view, the initial activation of lexical features is sense-neutral. In the case of homonyms, when the processor is faced with two separate meanings, a potentially costly selection process has to take place first before the ambiguity is resolved.

The idea of underspecification is compatible with the view that meaning assignment involves the initial activation of a lexical representation, followed by contextual specification and/or activation of more refined interpretations. This view is compatible with semantic accounts such as the Generative Lexicon approach (Pustejovsky 1995; see also Copestake and Briscoe (1995)), even though it is not clear what exactly is represented in an underspecified meaning. Arguments have been raised, however, against the idea of distinguishing between a semantic (sometimes called “core”) and a pragmatic (“non-core”) meaning (for an in-depth discussion, see e.g. Geeraerts (2010); cf. Chapter 11). To some degree, this discussion resembles current controversy over whether semantic memory is truly separate from general world knowledge (McRae and Jones 2013). At present we are unaware of conclusive experimental evidence favouring one account over the other and will therefore refrain from taking a position on this issue. It should also be noted that there might be gradations of polysemy. For example, Klepousniotou et al. (2008; see also Rabagliati and Snedeker 2013, discussed below) found differences in processing depending on how semantically overlapping the different senses of a polysemous word were. This issue clearly needs further research.

Finally, the decision to home in on a specific interpretation and the extent to which a language user homes in are likely to be affected by linguistic and extra-linguistic factors. For example, when a word is in focus position, it will attract more attention, possibly leading to a deeper level of interpretation (cf. the Moses illusion above). In addition, non-linguistic factors, such as the desire to critically evaluate information (e.g. as a reviewer), might also affect how quickly a more refined understanding will be achieved. On other occasions, the representation people build of a sentence may be much more superficial, though “good enough” for the task at hand (see Ferreira and Patson (2007)).
5 Combinatorial semantics

Semantic interpretation involves not only looking up word meanings but also combining them to arrive at the interpretation of a larger chunk of text. This follows from the principle of compositionality (“(t)he meaning of an expression is a function of the meanings of its parts and the way they’re syntactically combined”; Pylkkänen and McElree (2006), after Frege (1892); see also Chapter 24). Pylkkänen and McElree identified three basic compositional rules or operations that they believe should be part of any psycholinguistic account of interpretation. First, functional application (and semantic types) describes how basic argument structures are set up and interpreted (with assignment of thematic roles to arguments; see also Chapter 23). Second, predicate modification is a compositional rule that allows the intersective interpretation of adjective-noun constructions such as wise president. Third, predicate abstraction allows, among other things, the interpretation of relative clauses.

These operations are compatible with a large amount of psycholinguistic work. For example, with respect to functional application, evidence suggests that language users project thematic roles without having to wait for the argument to be actually expressed. Mauner et al. (1995; see also Carlson and Tanenhaus 1988) found that when a verb such as sink was used in a construction that implied an agent (The ship was sunk to collect a settlement from the insurance company), readers projected an implicit agent and processed the infinitival clause as fast as when the agent was expressed (The ship was sunk by the owners to collect . . .) and faster than when an implicit agent was not expected (The ship sank to collect . . .). This result shows that language users construct inferences on-line on the basis of a verb’s argument structure. Another example of functional application comes from the dissimilarity in the processing of arguments and adjuncts. People tend to read arguments faster than adjuncts (e.g. Clifton et al. (1991); Traxler (2008)) and have a preference for argument interpretation over adjunct interpretations of ambiguous phrases, indicating a preference to interpret new incoming information as being part of the evolving argument structure rather than being an optional sentential element (for in-depth discussions, see Frazier and Clifton (1996); Schütze and Gibson (1999); and Traxler and Tooley (2007)).

A challenge to strict compositionality comes from so-called coercive expressions, in which there is a semantic type mismatch blocking a straightforward application of basic compositional operations (see Jackendoff (2002); Pustejovsky (1995); Pylkkänen and McElree (2006)). One type of coercion, complement coercion, has attracted quite some interest from psycholinguists and has generated relatively consistent experimental results.

Complement coercion can occur when there is a semantic mismatch between the verb and its complement. For example, in (8), taken from McElree et al. (2001), the verb starting selects for an event complement. However, the complement the book denotes an entity and thus needs to be type-coerced into an event. In (9), the verb selects for an entity and this role is fulfilled by the complement the book.

(8) The author was starting the book in his house on the island.
(9) The author was writing the book in his house on the island.

Much evidence has revealed that interpreting complement coerced constructions such as (8) takes extra processing effort (and involves different brain processes) compared to a sentence such as (9), which has a straightforward, non-coerced interpretation (e.g. Kuperberg et al. (2010); Lapata et al. (2003); McElree et al. (2001); Pylkkänen and McElree (2007); Traxler et al. (2002); see also Frisson et al. (2011) for evidence from adjectival phrases such as difficult mountain).
To summarize the results from reading experiments, it has been shown that the coercion cost is unrelated to processing characteristics of the verb itself. For example, *start the fight*, in which the complement already refers to an event, does not generate a processing cost (Traxler et al. 2002), indicating that a verb such as *start* does not make processing of the complement more difficult by default, but only when paired with a non-eventive expression. It is also unrelated to differences in how acceptable language users judge these constructions (Traxler et al. 2005), in their plausibility (McElree et al. 2006), in cloze probability (Traxler et al. 2005), or in how often these types of verb co-occur with a specific complement in text (as measured by Latent Semantic Analysis co-occurrence frequencies; Traxler et al. (2002)). The cost also does not seem to be associated with having to generate or select an appropriate coerced event interpretation (Traxler et al. 2005), with having to select one interpretation among several possible ones or competition between different interpretations (Frisson and McElree 2008; Scheepers; Keller and Lapata 2008), nor with the relative frequency of a particular event interpretation (Frisson and McElree 2008). Rather, there is a fixed cost to coercion, which appears to reflect the time needed to set up a semantically enriched interpretation. In other words, to interpret a complement that does not have the required semantic type, extra semantic structure needs to be generated so that it can acquire the right type and this process is the locus of the cost.

A second type of coercion that has been investigated experimentally is aspectual coercion, which refers to the reinterpretation that occurs when the semantics of a punctual verb such as *jump* mismatches the semantics of an external modifier (e.g. *for hours*). However, since the psycho- and neuro-linguistic data on the processing of this type of coercion is less clear than for complement coercion, we will not discuss this construction further (see Brennan and Pylkkänen (2008); Pickering et al. (2006)).

6 (Im)plausibility

The idea of incrementality (see section 1 above) indicates that people are good at making sense of information as it unfolds. Results from homonym processing, which show immediate effects of relative meaning frequency, indicate that (at least some degree of) lexical processing occurs straight away. But when we process language, does the plausibility or likelihood of what we’re interpreting play a role and if so, how quickly does it affect processing? Alternatively, how good and quick are language users at noticing when something becomes strange?

Using eye-tracking, Rayner et al. (2004) examined how readers processed sentences containing either an inappropriate theme (using a pump to chop carrots), an unlikely theme (using an axe to chop carrots), or an acceptable theme (using a knife to chop carrots). Both the inappropriate and the unlikely condition led to processing difficulty but in different ways. Whereas the inappropriate condition led to immediate disruption in the eye movement record, the unlikely condition only showed (temporally and spatially) later disruption. In other words, reading about someone using an axe to chop carrots does not immediately strike one as totally implausible. These and similar (see Warren and McConnell (2007)) results make a case for a distinction between semantic knowledge and world knowledge: violations of semantic knowledge (e.g. a clash with a word’s selection restrictions) immediately disrupt processing, but mismatches with world knowledge, which express possible but very unlikely relationships, take somewhat longer to compute.
Warren et al. (2008; see also Ferguson and Sanford (2008)) hypothesized that if there were no distinction between the two types of memory and if the difference in processing merely reflects the degree with which the information clashes with real-world knowledge, then having a context which makes the implausible reading possible (e.g. a fairy-tale context) should eliminate the effect. However, the early processing cost was still observed, but the later processing cost, arguably reflecting higher-order, integrative and evaluative processing, was attenuated. Hence, these results suggest that some processing, such as managing selection restrictions, happens immediately and is impervious to higher-order information that could impact upon early interpretation. This view, however, is not accepted by all researchers and there is evidence more consistent with the view that processing at different levels – semantic and pragmatic – can occur at the same time. For example, Filik (2008; see also Matsuki et al. (2011)) found that implausibility effects (e.g. someone lifting up his car) could be overridden by an appropriate context (e.g. if that person is the Incredible Hulk).

More evidence in line with the unitary view comes from a series of brain imaging experiments, which point to a close link between semantic and encyclopaedic knowledge, with discourse violations eliciting comparable temporal and distributional effects to semantic violations (e.g. Hagoort et al. (2004); Nieuwland and van Berkum (2005); van Berkum et al. (1999); van Berkum et al. (2003)). For example, Nieuwland and van Berkum (2006) used sentences containing an animacy violation, such as (10) and (11):

(10) The girl comforted the clock.
(11) The peanut was in love.

In isolation, these sentences generated the expected N400 effect. However, when embedded in a larger context, this effect disappeared (e.g. a girl talking to a clock about its depression) or, even more strikingly, reversed the effect (e.g. in a story about an amorous peanut), such that the sentence containing the lexical-semantic violation became easier to process than one not containing the violation (the peanut was salted). (Note that this pattern is different from Urbach and Kutas’s (2010) finding discussed above. It’s not obvious why context can override an N400 effect in one case but not another.) Hagoort et al. (2004) compared sentences containing a semantic violation (12) or a pragmatic violation (13) (Dutch trains are uniquely yellow) to sentences containing no semantic or pragmatic violation (14).

(12) The Dutch trains are sour and very crowded.
(13) The Dutch trains are white and very crowded.
(14) The Dutch trains are yellow and very crowded.

They found that both types of violations resulted in an N400 effect, indicating that at the physiological level, there might not be a distinction between the integration of semantic and world knowledge information during sentence comprehension. It remains unclear, however, what an N400 effect exactly reflects in terms of on-line processing. In addition, Hagoort et al. did find a difference in oscillatory brain activity between the two types, though what exactly this implies is unclear. Clearly, the discussion about when exactly semantic and world knowledge comes into play has not been settled and it appears that experimentation using different methodologies, patient research (e.g. semantic dementia research: Graham et al. (2000)) and research in neighbouring fields (e.g. memory) may help resolve this issue.
7 Semantics in language production

As we have noted, the great majority of work on semantic processing relates to language comprehension rather than language production. Whereas researchers in the production of syntax appear to make use of similar linguistic models to those assumed by researchers in the comprehension of syntax, this does not appear to be the case in semantics. For example, Clark and his colleagues have conducted many investigations of how people select referring expressions. Language production involves many choices about whether to produce an utterance that makes all aspects of meaning explicit or whether to leave out certain components of content that can be inferred from some part of the context. Thus, responses to questions (e.g. *What time do you close?*) can be full (*We close at nine*) or elliptical (*At nine*; e.g. Clark (1979)). Similarly, referring expressions can be more or less detailed. For example, when speakers refer to an object that they have already successfully referred to, they can choose to produce a detailed referring expression (e.g. *looks like a person who’s ice-skating, except they’re sticking two arms out in front*), or – more often – choose a shorter but less detailed alternative (e.g. *the ice-skater*; Clark and Wilkes-Gibbs (1986); Krauss and Weinheimer (1964)). It turns out that people tend to produce shorter referring expressions as a conversation proceeds, so long as their interlocutors make it clear that they understand. In fact, many of these expressions can be seen as metonymic and the studies therefore provide evidence about how people may develop metonymy on the fly.

Research has also addressed the question of ambiguity avoidance. Thus, Ferreira et al. (2005) had participants name pictures of objects such as a baseball bat in the presence of a distractor (“foil”) object with a homophonous name, in this case a flying bat. Surprisingly, participants often mistakenly produced an ambiguous name, sometimes followed by a self-correction (e.g. *bat*. *no, baseball bat*). But when the foil was another instance of the same category (here, another baseball bat), they almost always avoided the ambiguity. Ferreira et al. also found that speakers often corrected themselves after producing an ambiguous expression, indicating that they detected the ambiguity “after the fact,” and that once corrected, they consistently avoided the ambiguity. This implies that ambiguity detection is much more effective after the expressions are produced compared to when they are about to be produced.

Rabagliati and Snedeker (2013) extended Ferreira et al.’s research and found that regular polysemes such as *chicken* (the animal and the meat) patterned with same-category descriptions (e.g. different instances of the baseball bat), while irregular polysemes such as *button* (shirt button and emergency button) patterned with true homonyms such as *bat*. This suggests that polysemes are represented and accessed differently from homophones and in fact are treated similarly to different exemplars of the same meaning of a homophone. For example, polysemes might access the same lemma, whereas homophones access different lemmas (Levelt et al. 1999). Moreover, people appear to represent regular polysemes together, but irregular polysemes separately (though the extent to which the different interpretations of the irregular polysemes were related to each other was not completely clear). These results suggest that speakers make use of fine-grained semantic distinctions during production (for supportive evidence from comprehension studies, see Klepousniotou et al. (2008)).

There are many other issues in language production that clearly relate to semantic processing. For example, many researchers have addressed the issue of the relationship between language and thought, such as in relation to space or time (e.g. see Casasanto (2008) and Chapter 9). Clearly the relevant aspects of language involve semantics. At the other extreme, a vast number of studies have used the picture-word interference paradigm, in which
participants name a picture while ignoring a printed or spoken word. Naming is generally slower when the word and picture are semantically related than when they are unrelated (e.g. Schriefers et al. (1990)). Finally, semantic relationships affect syntax, with participants being more likely to repeat syntactic structure (e.g. saying *the sheep that’s red*) after hearing an utterance with the same structure and a noun with related meaning (*the goat that’s red*) than a noun with unrelated meaning (*the book that’s red*; Cleland and Pickering (2003)). At this point, it appears that most of these studies are concerned with the relationship between semantic processing and other aspects of production rather than with the details of semantic representation itself.

### 8 Summary and conclusions

The study of semantic processing has for a long time played second fiddle to syntactic processing in psycholinguistic research. Over the last 10–15 years, this focus has shifted somewhat and more and more attention has gone to issues of lexical processing, semantic interpretation and their link with pragmatics. Lately, this focus has been extended into the field of language production and some researchers have started to look at how individual differences affect semantic processing (e.g. Hannon and Daneman (2004)).

When talking about semantic (and pragmatic) processing, the main focus is on what happens immediately: what information is used immediately and incrementally to build up a representation of what people are hearing or reading? How do comprehenders gain access to a word’s meaning? How and when does other information impact on this activation? By using sensitive measurements, we might be able to address these questions and arrive at a model of how bottom-up and top-down processes work together to give the language user a sense of understanding.

One way of making incremental processing faster is by predicting what will come next in the text. While predictability effects are robust and ubiquitous, they in themselves do not equate to semantic processing. For example, most experiments investigating a specific aspect of semantic processing control for predictability and consequently these effects hold independently from predictability effects. In addition, it remains unclear whether language users at every single moment in time calculate the odds of the next word occurring.

Research on words with multiple meanings (homonyms) and words with multiple senses (polysemes) generally points to different processing profiles for the two types of words. Since the processor, in the vast majority of cases, does not know whether a homonym or polyseme will follow next, the difference in processing could potentially be linked to differences in lexical representation. Research on combinatorial semantics, and specifically coercion, indicates that when an expression cannot be interpreted in a straightforward compositional manner, extra processing is required, which suggests that in most cases simple semantic composition is the default for interpretation. Research on plausibility effects, and specifically whether (pragmatic) context can overwrite (lexical) implausibility, is still far from conclusive, with some impressive examples of immediate pragmatic effects but also of some stubborn lexical-first effects. Finally, research on language production addresses some questions about representation considered in the much more extensive literature on comprehension (e.g. the nature of polysemy), but the literature has not so far addressed the relationship between semantic theory and the mechanisms of production in great detail.

Obviously, our overview has only scratched the surface of psycholinguistic semantic research and several topics have been woefully ignored. For example, research on figurative language interpretation has been extensive and has recently shifted more into the neurological
representation of literal and non-literal language (for overviews, see e.g. Coulson (2008); Gibbs and Colston (2006)). Similarly, work on perspective-taking during language processing, which can be related to Theory of Mind processing (e.g. Frisson and Wakefield (2012)) and on how individual (communicative) goals, situations and time constraints affect depth of processing (e.g. Ferreira and Patson (2007)) will need to be accommodated for in a psychologically realistic model of language comprehension.

Further reading


Traxler, M. J. 2012. Introduction to Psycholinguistics: Understanding Language Science. Sussex, UK: John Wiley and Sons Ltd. A textbook offering an excellent introduction to psycholinguistic research aimed at a diverse audience of psychologists, linguists, philosophers, computer scientists, etc.

Gibbs, R. W. Jr. and H. L. Colston 2012. Interpreting Figurative Meaning. New York: Cambridge University Press. This textbook brings together psycholinguistic and neurolinguistic research on figurative language and critically evaluates models of figurative language interpretation from different fields such as psychology, linguistics, and philosophy.


References


Steven Frisson and Martin J. Pickering


### Related topics

Chapter 2, Internalist semantics; Chapter 8, Embodiment, simulation and meaning; Chapter 9, Linguistic relativity; Chapter 10, Semantics and pragmatics; Chapter 11, Contextual adjustment of meaning; Chapter 13, Sense individuation.