Eye movements and semantic composition

Article · January 2004

4 authors, including:

Martin J Pickering
The University of Edinburgh
319 PUBLICATIONS   **20,680** CITATIONS

SEE PROFILE

Steven Frisson
University of Birmingham
65 PUBLICATIONS   **1,653** CITATIONS

SEE PROFILE

Brian Mcelree
New York University
49 PUBLICATIONS   **2,878** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Coordinating utterances during turn-taking View project

ExLing 2021: 12th International Conference of Experimental Linguistics View project
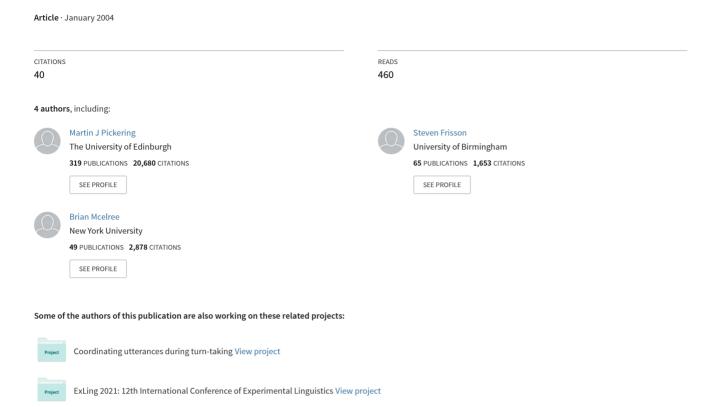
Eye movements and semantic composition

Martin J. Pickering

Department of Psychology

University of Edinburgh

Steven Frisson

Department of Psychology

University of Massachusetts, Amherst

Brian McElree

Department of Psychology

New York University

Matthew J. Traxler

Department of Psychology

University of California, Davis

Short Title: Eye movements and semantic composition

Address Correspondence to:

Martin Pickering

Department of Psychology

University of Edinburgh

7 George Square

Edinburgh EH8 9LW

U.K.

e-mail: martin.pickering@ed.ac.uk

phone: +44 (0) 141 330 6847

fax: +44 (0) 141 330 3461

1

Most research on language comprehension has focused on what we might consider the two ends of the process. One tradition is concerned with the "precursors" to interpretation – how people recognize words, how they decide which of two unrelated meanings of a word is appropriate, and how they assign a syntactic analysis. The other tradition is predominantly concerned with the way in which interpretations of whole sentences are combined – how people determine the anaphoric links between sentences, what kinds of inferences they draw, and how they determine what texts are "about". But between these two extremes, there is another question (Frazier, 1999): how do people come up with the appropriate interpretation for complex expressions, such as phrases and sentences? In this chapter, we seek to explore semantic composition by monitoring eye movements during reading.

We concentrate on two types of expression that we have considered in recent work. The first concerns the interpretation of <u>sense</u> ambiguities, where a word has two (or more) related interpretations. Although almost every word is ambiguous at the level of sense, the vast majority of psycholinguistic research on lexical ambiguity has concentrated on the rarer instances of <u>meaning</u> ambiguity, where a word (e.g., <u>bank</u>, <u>coach</u>) has unrelated interpretations (e.g., Swinney, 1979; Rayner & Duffy, 1986). Here, the main focus is on what has traditionally been termed metonymy: <u>Dickens</u> can refer literally to the man or metonymically to his writings, and <u>Vietnam</u> can refer literally to the country or metonymically to the Vietnam War. The second type of expression that we are concerned with is illustrated by <u>began the book</u>, which can mean "began reading the book", "began writing the book", or indeed other things (Jackendoff, 1997; Pustejovsky, 1995). Such ambiguities appear to arise when the semantic requirements of the verb are not met by the complement; here, because a verb like <u>began</u> requires a complement that denotes an event, but when the actual complement is a noun phrase that denotes an entity. The processor resolves this problem by inserting relevant semantic content (reading, writing, or whatever) via what has been termed "enriched composition". An interesting theoretical question is the extent to which the resolution of sense ambiguity and the application of enriched composition can be related.

Many aspects of language comprehension take place without appreciable delay, and there is considerable evidence for what has often been termed "incrementality" – processing each new piece of information as soon as it is encountered (e.g., Marslen-Wilson, 1973; see Pickering, 1999, for a review). For

example, it might have been the case that people did not interpret each word in a sentence incrementally, or interpreted words incrementally but only constructed a syntactic analysis at the end of the sentence, or only interpreted anaphora at the end of the sentence. But we now know that people do not greatly delay such aspects of processing.

During language comprehension, any effect may in principle be immediate or delayed. Although there is certainly a continuum, some effects clearly occur without delay and some clearly can be greatly delayed. For example, people are rapidly disrupted by a badly misspelt word, a non-word, or a word written using inappropriate capitalization, and they take longer reading a rare word than a frequent word (e.g., Rayner & Pollatsek, 1989). Hence, the aspects of processing that underlie these effects (e.g., lexical access) must be rapid as well. In contrast, people often take a long time to realize that a difficult argument was illogical or based on false premises.

Just and Carpenter (1980) proposed the immediacy hypothesis, whereby readers try to interpret each content word as it is encountered. This means that they encode the word and select an interpretation (on the basis of frequency, context, or whatever), assign it to its referent, and determine its status in the sentence and discourse. Though these claims are vague in many respects, what they appear to amount to is that the reader does not delay in performing lexical access, disambiguates the word with respect to context, constructs a syntactic representation that incorporates the word, and produces a semantic representation that is assessed with respect to background knowledge. All of this processing occurs before the reader starts processing the next word (assuming of course that the reader is successful). In this chapter, we suggest that immediacy may be compromised for some aspects of semantic composition. The point is illustrated this with reference to sense ambiguity resolution and semantic composition.

Determining the time-course of processing using eye movements

When a reader encounters a rare word, for example, there may be immediate signs of processing difficulty. For this to happen, readers must be bound by two assumptions: immediacy, as already discussed; and what has been termed the eye-mind assumption, which is the assumption that people look at the thing they are thinking about. Jointly, the assumptions entail that if a reader is looking at a particular word, for instance, then

she is processing that word. If an effect linked to a word occurs while a reader fixates on the word, then we can be certain that the relevant aspect of processing occurred immediately (in accord with the immediacy hypothesis) and that the eye-mind assumption holds. So for instance, if people look at rare words for longer than frequent words (Just & Carpenter, 1980), then we can infer that they perform lexical access whilst fixating that word, and that rare word are more difficult to identify than frequent words.

However, if an effect is delayed, or spills over onto the next word (Ehrlich & Rayner, 1981), then we cannot be certain whether immediacy does not hold (i.e., whether they delay processing) or whether that particular effect does not obey the eye-mind assumption. For example, some aspects of higher level processing may require the reader to "think about" a discourse model, and fixating a particular word for a longer time does not help with that process. Another possibility is that a reader may encounter difficulty at a particular point in a text, but decide that it makes sense to continue moving forward as normal, perhaps to see whether future information may resolve the problem. For example, in work discussed below, Frisson and Pickering (1999) found delayed difficulty with "A lot of Americans protested during Finland …". Although readers may have immediately realized that they could not think of a relevant event associated with Finland, they may have kept moving forward to see if future information resolved their problem, or because they realized that finding an event associated with Finland involved searching long-term memory and did not require any attention to a particular point in the text.

The implication is that it is not possible to use a delay in observed disruptions of the eye movement record to unequivocally reject the immediacy hypothesis, as it is possible that, instead, the eye-mind assumption does not hold for that aspect of processing. Nevertheless, evidence that an effect does not show up immediately, when there is evidence that comparable effects do show up immediately, suggests that the cognitive processes underlying the effect are at least different from those involved in other processes that reveal immediate effects. More generally, differences in the time-course and pattern of effects are likely to provide grounds for differentiating processes.

Recent experimental work in language comprehension using eye-movements has employed many different measurements to determine the existence, locus, and time-course of processing difficulty (see Liversedge, Paterson, & Pickering, 1998;

Rayner, 1998; Rayner et al., 1989). It is first necessary to define the region of interest, which almost always involves one or more words, because it is standardly assumed that the word is the basic unit of analysis in sentence comprehension (rather than the character or morpheme, for instance). Analyses are almost entirely based on time spent in a region rather than length of saccade, as these measurements appear to be most sensitive to cognitive processing. An exception is analyses based on proportion of regressive eye movements. (Time spent during a saccade is usually excluded, but it probably does not matter greatly whether it is counted or not.) The "earliest" measure is first fixation, which is the time of the first fixation on a word. This appears to be sensitive to some cognitive processes (e.g., word frequency) and some gross syntactic anomalies. A slightly later measure that is probably more revealing for most issues in sentence comprehension is first-pass time, which is normally defined as the sum of all fixations beginning with the reader's first fixation in a region until the reader's gaze leaves the region (on one-word regions, first-pass time is equivalent to gaze duration; e.g., Rayner & Duffy, 1986). For both these measures, most researchers exclude trials on which readers "skip" the region during initial processing.

It is important to stress that first-pass effects cannot simply be equated with effects occurring during early phases of cognitive processing. This is because we do not typically know how to uniquely map cognitive events on to the different eye-movement measures, even when we have good grounds for believing that the eye-mind assumption generally holds. For example, even if both theoretical and empirical considerations strongly suggest that processing is split into two stages (initial analysis and reanalysis, lexical access and contextual integration, etc.), we cannot be certain that certain measures (e.g., first-pass time) exclusively measure the initial stage. Clearly, we would need a very specific hypothesis (backed up by relevant evidence) to "link" eye movement patterns with underlying cognitive events.

An additional and more practical issue is that the linking is affected by the length of the region of interest. A first-pass effect for a "long" region (say, two or three words) will be clearly different from one that occurs on a single word. Another problem with drawing conclusions about long regions is that first-pass time finishes when the eye leaves the region in either direction, and so an regression from the first word of a long region back to an earlier region leads to a short first-pass time, although such a regression is presumably indicative of immediate processing

difficulty. Hence, our preference has always been to define one word critical regions where possible. Under such conditions, first-pass time, like first-fixation time, is spatially well-localized.

Other reading time measures capture somewhat later effects. Traditionally, these have included total time, defined as the sum of all fixations on a word, and second-pass time, which can be defined in slightly different ways. In our research, we have tended to define second-pass time as the time spent in a region after leaving the region (or after an initial "skip" of the region). However, just as it is dangerous to assume that first-pass time measures events that take place during initial cognitive processing, it is dangerous to assume that second-pass time measures reanalysis. A similar point holds for other definitions of second-pass (or re-reading) time, such as time spent on a region following a regressive saccade.

Further measures have been developed and tested during recent years. The most important "new" measure, which often seems to capture interesting effects, is most commonly called regression-path time, though other names exist (Brysbaert & Mitchell, 1996; Konieczny, Hemforth, Scheepers, & Strube, 1997; Liversedge, 1994; Traxler, Bybee, & Pickering, 1997). This includes all fixations from the first within a region until the first to the right of a region (again, with "skips" generally excluded). Hence, it includes fixations outside the target region when the reader has regressed to the left prior to fixating to the right of the region. It is striking how frequently this measure has been used in recent papers, suggesting that it is a particularly informative measure. One reason it is interesting is because we know that the reader has not looked at anything after the critical word. Regression-path time is sometimes interpreted as the time necessary for a reader to process the text to a sufficient degree that she is prepared to input new information (and appears to correspond best to what a reader can do during cumulative self-paced reading). However, note that some fixations included in the measure may be quite "late," because the reader may repeatedly fixate the beginning of the sentence after fixating the target word. Consequently, if first-pass time does not show any difference but regression-path time does, it must be that the effect is due to those trials on which a regression occurred from the region in question.

Finally, many papers report proportions (or numbers) of regressions. It is crucial to distinguish first-pass regressions from later measures of regressions. A first-pass regression from a region is a leftward movement from that region following

a first-pass fixation in that region. Clearly an effect on first-pass regressions is an early effect, because it demonstrates that readers have made a different decision for one condition over another during the time that they first spend in that region. In contrast, a measure of regressions into a region (i.e., back from a later region) is very akin to second-pass time and does not reflect early processing.

What do these different measures tell us about language comprehension? The answer is of course that they may tell us a lot of different things. There is no single measure that should be preferred, and they are enormously affected by the characteristics of the regions chosen. If effects emerge on first-fixation or first-pass time for a single word of roughly normal length, we can be fairly confident that they emerged during the earliest stages of processing that eye-tracking is likely to detect (discounting any preview effects of course). First-pass regressions would also be informative of early effects, if they occur on the target region. Very early spill-over effects (e.g., first fixation after the target word) must also be fairly early. Apart from this, it is not clear what can be said with certainty without a theory of how eye movements and cognitive processing interact. As noted above, it is very unlikely that that there is a straightforward relationship between, on the one hand, first-pass and first-fixation measures and initial processing, and, on the other, between later measures and later processing. For the most part, we consider whether effects occur on "early" measures of processing or not, on the assumption that the combination of the immediacy and eye-mind hypotheses predicts that effects of linguistic processing should occur as soon as the relevant word is fixated.

Sense resolution

As noted above, most work on lexical ambiguity resolution has been concerned with words with unrelated meanings (i.e., homonyms). For example, the two meanings of bank (roughly, repository for money, and side of a river) are largely unrelated. There has been a great deal of research on the "on-line" resolution of homonyms, using priming (Seidenberg, Tanenhaus, Leiman, & Beinkowski, 1982; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979), eye-tracking (Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989), or a combination of the two (Sereno, 1995). This work has been greatly influenced by the "modularity" debate (Fodor, 1983) and has concentrated on the questions of whether all meanings are

activated irrespective of context, and how and when the processor selects one meaning. Meaning resolution can be thought of as occurring before an interpretation is assigned to complex expressions. Context basically serves to select one meaning of a homonym, and this meaning is subsequently integrated with the context.

There has, in contrast, been very little work on the processing of words with related senses, but it is clear that sense resolution is very different from meaning resolution. Most words have only one or two meanings. In contrast, almost all words have very many senses, and it can be very hard to enumerate them. For example, newspaper can refer to the institution, the content of the newspaper over a number of edition, the day's edition, or a particular copy (e.g., Copestake & Briscoe, 1995), window can refer to the glass (broke the window) or to the aperture (climbed through the window), and even a proper name like Julius Caesar can refer to the man, a picture or statue of him, or even a speech by him, given appropriate context. It is also far from clear that either producers or comprehenders always determine exactly which sense is appropriate (e.g., Sanford & Sturt, 2002), and conversationalists often only resolve what they are talking about to a depth necessary for current purposes (Clark & Wilkes-Gibbs, 1986). However, sense resolution obviously can occur: For example, "I tore the newspaper" requires that newspaper refers to a particular copy. Although it is theoretically possible that sense resolution occurs like meaning resolution, a much more likely alternative is that it takes place as part of the process of assigning an interpretation to a complex expression.

There have been some attempts to study the representation of sense ambiguity, with Klein and Murphy (2001, 2002) arguing that different senses are represented in a largely independent manner, so that the distinction between sense and meaning representation is a matter of degree rather than kind. In contrast, Rodd, Gaskell, and Marslen-Wilson (2002) proposed clear differences between sense and meaning ambiguity with respect to their effects on comprehension time.

There have been few studies on the processing of sense ambiguities in sentence contexts. One important exception is Frazier and Rayner (1990), who used eye-tracking to investigate the processing of nouns with multiple senses (e.g., novel) and multiple meanings (e.g., ball), and compared them to unambiguous nouns like door. Their results showed that words with multiple meanings were processed differently from words with multiple senses, which behaved much more like unambiguous words. For example, when disambiguating information followed the

ambiguous noun, nouns disambiguated to the subordinate meaning caused more difficulty than nouns disambiguated to the dominant meaning. Apparently people generally adopted the dominant meaning on, or soon after, encountering the word. However, words with multiple senses, which were disambiguated towards either the concrete or the abstract interpretation, did not show processing difficulties. Thus, they concluded that people did not perform immediate sense disambiguation in the absence of prior disambiguating context and that an abstract sense could be processed as fast as a concrete sense. Frazier and Rayner proposed a "minimal semantic commitment" model in which a single semantic value is only assigned immediately when two interpretations are incompatible.

To address the time-course of sense ambiguity resolution, Frisson and Pickering (1999) turned to the consideration of metonymy. In metonymy, a salient aspect of an entity is used to refer to the entity as a whole or to some other part of the entity. In the following examples of metonymy, definitions of the relationship between the literal and metonymic interpretations are provided in parentheses:

1a. I read Dickens whenever I can (producer-for-product)
1b. The wings took off from the runway (part-for-whole or synecdoche)
1c. Belgium will win the European cup (whole-for-part)
1d. The blasphemous woman had to answer to the convent (place-for-institution)
1e. They protested during Vietnam (place-for-event)

Frisson and Pickering argued that metonymy was ideal for psycholinguistic investigation using eye-tracking because it is generally localized to a single word. Hence, we can precisely determine the time-course of metonymic processing. This contrasts with, for example, idioms and most metaphors, which are generally "dragged out" over a number of words. In such cases, it is often difficult to determine the precise point at which it becomes clear that the expression is not being used literally (consider, for example, That's the way the cookie crumbles).

Frisson and Pickering (1999) considered the processing of place-for-institution metonyms like convent in (1d) and place-for-institution metonyms like Vietnam in (1e). In Experiment 1, they contrasted literal versus metonymic uses of convent with a control word stadium, that does not have a relevant metonymic use:

2a. These two businessmen tried to purchase the convent at the end of last April, which upset quite a lot of people.

2b. That blasphemous woman had to answer to the convent at the end of last March, but did not get a lot of support.

2c. These two businessmen tried to purchase the stadium at the end of last April, which upset quite a lot of people.

2d. That blasphemous woman had to answer to the stadium at the end of last March, but did not get a lot of support.

Readers experienced no difficulty with <u>convent</u> in either the metonymic condition (2b) or the literal condition (2a). This suggests that the processor could rapidly access both senses. Although it is possible that the two senses were computed in parallel, with context then selecting between them, the lack of any correlation between sense frequency and processing difficulty makes this account unlikely, and so Frisson and Pickering proposed that readers initially accessed an underspecified representation that did not distinguish between the literal and metonymic senses (see also Frisson & Pickering, 2001). In contrast, readers experienced immediate difficulty with <u>stadium</u> in (2d), with the first-pass time on <u>stadium</u> being longer in (2d) than (2c). This effect is particularly early, and suggests that effects of semantic incongruity can occur as soon as a word is fixated, in accord with Just and Carpenter (1980).

In Experiment 2, they contrasted literal versus metonymic uses of words like <u>Vietnam</u> with control words for places that were not associated with any event:

3a. During my trip, I hitchhiked around Vietnam, but in the end I decided to rent a car for a couple of days.

3b. A lot of Americans protested during Vietnam, but in the end this did not alter the president's decision.

3c. During my trip, I hitchhiked around Finland, but in the end I decided to rent a car for a couple of days.

3d. A lot of Americans protested during Finland, but in the end this did not alter the president's decision.

Readers experienced no difficulty with (3a) or (3b), suggesting that they could access both senses without difficulty. They did experience difficulty with (3d), but the

difficulty was delayed quite considerably (it appeared on regressions from an end-of-line region corresponding to I decided to/this did not for this example, and on late measures). A pre-test indicated that (3d) was implausible (in comparison to 3c), though considerably less implausible than (2d). However, almost no-one associated any event with Finland, so the manipulation could not have been stronger for this condition. Thus it appears that the delayed difficulty with (3d) is a genuine demonstration that a particular kind of incongruity effect is delayed.

These experiments suggest that established senses are interpreted straightforwardly, and do not lead to any processing difficulty even if they are figurative. Pickering and Frisson (2003) considered whether it was also possible to make a novel sense familiar using appropriate context. They employed the producer-for-product rule (roughly, if X refers to a producer, then X also refers to that producer's characteristic product).

4a. Not so long before she died, my great-grandmother met Dickens in the street. I heard that she often read Dickens when she had the time.

4b. My great-grandmother has all the novels written by Dickens in her library. I heard that she often read Dickens when she had the time.

4c. My great-grandmother confessed that she once kissed Dickens on the cheek. I heard that she often met Dickens when she had the time.

4d. Not so long before she died, my great-grandmother met Needham in the street. I heard that she often read Needham when she had the time.

4e. My great-grandmother has all the novels written by Needham in her library. I heard that she often read Needham when she had the time.

4f. My great-grandmother confessed that she once kissed Needham on the cheek. I heard that she often met Needham when she had the time.

Texts (4a-c) employed a familiar producer, whereas (4d-f) employed an unfamiliar one (i.e., Needham wasn't identified as an writer). In the second sentence, (4c & f) used the name to refer to a person, whereas (4a, b, d, & e) used the name to refer to the product. We made two predictions. First, we predicted that Dickens would be easier to read than Needham, because it is a more familiar name (names were matched for length), on both first and second mentions of the names. Name familiarity is presumably closely related to frequency (more familiar names being more frequent),

and frequency effects occur during early processing (Just & Carpenter, 1980; Rayner & Duffy, 1986). Therefore we predicted effects of name familiarity should occur during early eye-tracking measures (i.e., first-pass and first-fixation time), and they did.

Second, we predicted difficulty when a name was used to refer to a product, when readers weren't aware that it referred to a producer. Clearly, no difficulty is predicted with the literal uses of either <u>Needham</u> or <u>Dickens</u> (i.e., with 4c or 4f), or with the metonymic use of <u>Dickens</u> (i.e., with 4a or 4b). In contrast, (4d) should be difficult because the context sentence does not introduce Needham as a writer, so the second reference to Needham (in <u>read Needham</u>) should be difficult in just the same way that the reference to <u>Finland</u> in (3d) was. Again, we found this pattern of results, with effects occurring on late measures (specifically, second-pass time on and just before the name).

The most interesting condition is (4e), where the context sentence introduces Needham as a producer. We found that the target sentence was easier in (4e) than (4d). In fact, there was only weak evidence of any difficulty with (4e) at all (i.e., in comparison to 4f). This demonstrates that a single mention of a word in the appropriate sense can be enough for novel sense learning. But most strikingly, the first mention of Needham in (4e) is literal. The metonymic sense of Needham has never been encountered before. Hence, the new sense is learned not by repeated exposure but rather by application of the producer-for-product rule.

These three experiments provide evidence about the applicability of Just and Carpenter's hypotheses. We found immediate effects of the difference between <u>Needham</u> and <u>Dickens</u> and the difficulty of <u>answer to the stadium</u>, and delayed effects of <u>during Finland</u> and <u>read Needham</u>. The early name familiarity effect demonstrates the sensitivity of our experiments (i.e., it shows that we can detect early signs of difficulty). We propose that difficulty is delayed for <u>during Finland</u> and <u>read Needham</u> because they involve unsuccessful searches of their general knowledge. In other words, readers initially encounter these phrases and try to determine whether something relevant happened in Finland and whether they have heard of a writer called Needham. In other words, they start to explore their general knowledge at this point. This search is sufficiently "detached" from the word <u>Finland</u> or <u>Needham</u> itself that the reader does not continue to fixate the word. Only after it becomes clear that the search is likely to be unsuccessful do readers refixate the critical word. Whether

this account is right in detail or not, the results are incompatible with the immediacy hypothesis, because the effects are localized but during second pass.

So why does immediate difficulty emerge with <u>answer to the stadium</u>? Intuitively, it appears to be that readers do not search general knowledge to determine if a stadium can be addressed. Instead, it appears to be part of the meaning of the word <u>stadium</u> that it is an inanimate object, and it is not plausible to address an inanimate object. Of course, it is possible to use stadium metonymically to refer to the people in the stadium, but this sense is not sufficiently familiar (without contextual support) for it to be accessed automatically. In other words, the difficulty with <u>answer to the stadium</u> appears to be due to the lexical representations of the words, whereas the difficulty with <u>during Finland</u> or <u>read Needham</u> appears to be one of reference. In any case, the results point to the complexity of determining which effects should occur on early processing measures and which should not. It is probably only because the regions are sufficiently small and well-defined that such distinctions can be drawn.

Finally, Pickering and Frisson (2001) considered the processing of verbs with multiple senses where the dominant sense was considerably more frequent than the subordinate sense. Pre-tests determined that the senses were closely related.

5a. After the capture of the village, we <u>disarmed</u> almost every rebel and sent them to prison for a very long time. (Supportive preceding context, dominant sense):
5b. With his wit and humour, the speaker <u>disarmed</u> almost every critic who was opposed to spending more money on art. (Supportive preceding context, subordinate sense):
5c. Mr Graham is quite certain that they <u>disarmed</u> almost every rebel and sent them to prison for a very long time. (Neutral preceding context, dominant sense):
5d. Mr Graham is quite certain that they <u>disarmed</u> almost every critic who was opposed to spending more money on art. (Neutral preceding context, subordinate sense):

The most important finding was that processing difficulty with the subordinate sense relative to the dominant sense did not emerge until much after the critical verb was first encountered. Similarly, prior disambiguating context reduced processing difficulty, but again this difference did not emerge until later processing. Again, these

13

results are incompatible with immediacy. Interestingly, another experiment using verbs with two meanings (e.g., ruled) also showed delayed effects of meaning dominance, suggesting that verbs may be processed differently from nouns (where the subordinate bias effect occurs during initial processing; Duffy et al., 1988). However, verbs with multiple meanings and verbs with multiple senses were not processed in the same way: Context had relatively rapid effects on meaning resolution but not on sense resolution.

Semantic composition

It is sometimes possible to derive the appropriate interpretation for a complex expression by simply combining key semantic properties of the individual words according to their syntactic position in the sentence (e.g., Jackendoff, 1997, 2002). In such circumstances, semantic properties retrieved from lexical representations and grammatical constraints associated with syntactic representations will uniquely determine the interpretation of the expression. Accordingly, the compositional mechanism—the critical interface between lexical and syntactic processing on one hand and discourse and text comprehension on the other—might merely consist of rules or principles for recursively combining semantic properties.

However, many common and seemingly simple expressions appear to require a richer form of composition (Jackendoff, 1997, 2002; Pustejovsky, 1991, 1995). An example of such an expression is The boy began the book. The verb begin (like start, enjoy, etc.) requires a complement with an event meaning (e.g., the fight, the assignment, the movie). But the default interpretation of the book is an entity. In order to satisfy the semantic requirements of begin, it must be construed as an event. Typically, the complement is interpreted as an event by implicitly generating an activity that is commonly associated with the complement noun and compatible with the agent. Thus, the boy began the book is typically interpreted as the boy began reading the book, whereas the author began the book is typically interpreted as the author began writing the book (McElree, Pickering, Traxler, Seely, & Jackendoff, 2001; Traxler, Pickering, & McElree, 2002). To do this, the complement is type-shifted from an entity to an event interpretation, and unexpressed semantic content (e.g., reading, writing) must be introduced.

Piñango, Zurif, and Jackendoff (1999) presented some evidence for difficulty with enriched composition. Using a cross-modal lexical decision task, they found greater processing load after <u>The insect hopped effortlessly until …</u> versus <u>The insect glided effortlessly until …</u>. They argued that people interpreted <u>hopped</u> as referring to a point-action event with an intrinsic beginning and end, and were therefore forced to shift its aspectual form from point-action to repeated activity when they encountered <u>until</u>, whereas no such type-shift was necessary following <u>glided</u>. It is possible to interpret this difficulty as a kind of garden-path, where <u>hopped</u> is initially assigned the wrong aspectual form.

We now have good evidence for difficulty with enriched composition from a quite different source, and also some suggestion that such difficulty is typically somewhat delayed. In a self-paced reading study, McElree et al. (2001) contrasted the following:

6a. The author was starting the book in his house on the island.
6b. The author was writing the book in his house on the island.
6c. The author was reading the book in his house on the island.

Sentence (6a) involves enriched composition, whereas (6b & c) do not. The verb in (6b) was employed in the most common interpretation of (6a) (i.e., started writing the book), and the verb in (6c) was employed in the next most common interpretation of (6a) (i.e., started reading the book), as assessed by a fill-in-the-blank pre-test. McElree et al. found difficulty with (6a) at <u>book</u> and <u>in</u>, suggesting that enriched composition caused a noticeable processing cost.

Traxler et al. (2002) ran an eye-tracking version of this experiment, and also found difficulty with sentences similar to (6a) in comparison to controls like (6b & c). In this study, difficulty did not emerge on first-pass time on <u>the book</u>. Although there was some sign of a first-pass regressions difference on <u>the book</u>, there was no effect on regression-path time. The clearest effects occurred on regression-path time from the words immediately following <u>the book</u>, and on other late measures. Because readers would have realized the need for enriched composition by <u>book</u>, these results again provide evidence against Just and Carpenter's assumptions and suggest that some semantic processing has delayed effects on the eye-movement record.

McElree et al. (2001) argued that an event-taking verb with an entity complement causes difficulty. However, it is possible that verbs like starting are difficult for some unrelated reason. To distinguish these possibilities, Traxler et al. contrasted event complements (e.g., the fight) with entity complements (e.g., the puzzle), and verbs that require an entity complement (e.g., started) with verbs that are compatible with either an entity or an event complement (e.g., saw):

7a. The boy started the fight after school today.

7b. The boy saw the fight after school today.

7c. The boy started the puzzle after school today.

7d. The boy saw the puzzle after school today.

In accord with our predictions, started the puzzle caused processing difficulty, but started the fight did not. Overall, started the fight was processed similarly to saw the fight, but started the puzzle caused more difficulty than started the fight or saw the puzzle. This demonstrates that the cause of difficulty is the combination of entity-taking verb and event noun. In this experiment, difficulty was clearly delayed, with effects emerging on regressions from a post-target region (here, after school), and on second-pass and total-time measures on the noun phrase itself. A self-paced reading replication of this experiment showed this pattern of effects on after but not on puzzle/fight.

Although we can now conclude that the specific combination of event-taking verb with entity complement causes processing difficulty, there are different possible explanations for the difficulty. It could merely be due to anomaly detection, with the "clash" between the semantic type of the required and the actual complement causing difficulty. This would therefore be analogous to many garden-path effects in parsing, for instance when an intransitive verb is followed by a noun phrase (e.g., Mitchell, 1987). Whereas anomaly detection may contribute to difficulty, it leaves unanswered the question of how the processor actually obtains the enriched interpretation for the expression.

Alternatively, difficulty may be due to ambiguity. Almost all cases of enriched composition are formally ambiguous, so that start the puzzle is probably most likely to mean "started solving the puzzle" but could mean "started making the puzzle," and can clearly mean other things in appropriate contexts (e.g., "started

16

painting the puzzle" in a context about an artist who was painting things). In some cases, ambiguity is a source of processing difficulty. For example, it is more difficult to process nouns that have two meanings of roughly equal frequency (balanced nouns) than otherwise comparable unambiguous words in a context that does not provide disambiguation (Rayner & Duffy, 1986). Most likely, readers have difficulty deciding which interpretation to select for such ambiguous words. Could enriched composition be similar, with the difficulty being due to the process of selecting the appropriate interpretation?

There are a number of reasons why this explanation is unlikely to be correct. Ambiguity effects are quite limited, in that they appear to occur for balanced nouns with multiple meanings in neutral contexts alone. They may well not occur for syntactic ambiguity (Traxler, Pickering, & Clifton, 1998; Van Gompel, Pickering, & Traxler, 2001) or for words with multiple senses (Frisson & Pickering, 1999), and require the two meanings to be of similar frequency (Rayner & Duffy, 1986). Duffy et al. (1988) found that a context that supported one interpretation removed the difficulty with the interpretation of balanced words. The ambiguity effects also appear to occur immediately, suggesting that the effect occurs during the process of selecting a meaning. In contrast, our effects are often somewhat delayed. More importantly, all our experiments involve highly biased items, with one interpretation (e.g., "the author started writing the book") being much preferred to alternatives (as determined by fill-in-the-blank pre-tests).

Additionally, Traxler, McElree, Williams, and Pickering (2003) found that the difficulty is not attenuated by a context that provides strong support for the preferred interpretation. Participants read two-sentence texts, in which the second sentence described an event using a coercing verb (8a, 8b) or a verb that corresponded to the coerced interpretation (8c, 8d) (i.e., coercing vs. control target). The first sentence either mentioned this event (8a, 8c) or did not (8b, 8d) (i.e., relevant vs. neutral context).

8a. The contractor had been building in the suburbs. That spring, he began a condominium next to the shopping center.
8b. The contractor had been looking for new jobs. That spring, he began a condominium next to the shopping center.

8c. The contractor had been building in the suburbs. That spring, he built a condominium next to the shopping center.
8d. The contractor had been looking for new jobs. That spring, he built a condominium next to the shopping center.

In the neutral context conditions, we predicted difficulty for the coerced target sentence (8b) versus the control target sentence (8d). If a context mentioning the coerced event affected processing of the target sentence, we would not expect a similar difference in the relevant context conditions. In other words, (8a) should be no more difficult than (8c). In fact, we found the same difference between (8b) and (8d) as between (8a) and (8c). In other words, removing the ambiguity did not eliminate the coercion cost. More specifically, the experiment showed the coercion cost remained when there was a greatly preferred interpretation for the coerced expression and when context supported this interpretation by mentioning the event (building). Again, these effects primarily emerged on later measure such as total time (rather than, for example, first-pass time on condominium). We can therefore rule out an explanation of coercion cost in terms of ambiguity. The results also suggest that coercion does not appear to be a kind of bridging inference (Clark & Haviland, 1977), because the context should have provided the bridge. Instead, it appears to be due to language-internal semantic operations.

We might, however, predict that it would be possible to prime these semantic operations by repeating either the exact operation or its interpretation. Traxler et al. (2003) had participants read coercing targets (9a, 9b) or control targets (9c, 9d) following coercing contexts (9a, 9c) or control contexts (9b, 9d). The preferred interpretation of the coerced contexts and targets instantiated the same interpretation as the control contexts and targets, and hence the verb (started or read) was used transitively with the same entity serving as the object as in the target sentence. Most importantly, the context and target sentences involved the same interpretation of the complement:

9a. The student started a book in his dorm room. Before he started the book about the opium trade, he checked his e-mail.
9b. The student read a book in his dorm room. Before he started the book about the opium trade, he checked his e-mail.

18

9c. The student started a book in his dorm room. Before he read the book about the opium trade, he checked his e-mail.

9d. The student read a book in his dorm room. Before he read the book about the opium trade, he checked his e-mail.

When participants read the coerced target, they had already interpreted the complement in the same way in the context sentence. In (9a), they had done this by performing the same coercion operation, and therefore constructing the same complement to the coercing verb. In (9b), they had arrived at the same interpretation for the verb phrase in the context sentence via normal composition.

In fact, both forms of context removed the difficulty with the coerced target. More specifically, Traxler et al. (2003) found difficulty reading the coerced context sentence versus the control context sentence (thereby confirming the basic pattern of difficulty with coercion) but no difficulty reading the coerced target sentence versus the control target sentence in either condition (and moreover, there was an interaction between context and target sentences). As before, the clearest effects emerged on total time, though this experiment did find evidence for an effect on the first noun region (i.e., a book in 9) on regression-path time. In other words, the construction of the interpretation "read a book," whether via enriched or normal composition, removed the difficulty with assigning that interpretation to started the book in the following sentence. A further experiment found similar effects when the full noun phrase the book was replaced with an anaphor (it) in the target sentence.

These experiments suggest that the difficulty of performing coercion is attenuated by recent processing of the relevant event structure, either directly (e.g., read a book) or following coercion (e.g., started a book). In contrast, a relevant context that simply mentions the event (building in 8) is not effective. Hence, simply facilitating the "bridging" inference is not enough to remove the difficulty. This suggests that coercion is a language-internal operation whose difficulty cannot be removed simply by making the coercion more compatible with general context. However, it is possible to prime the actual mechanisms involved in the coercion.

Does type-shifting cause difficulty?

At this point, we have evidence for difficulty with the semantic operations involved in enriched composition, but evidence for no difficulty with the semantic operations in understanding familiar metonymy. Indeed, the coercion studies find difficulty with the computation of a complement with an event interpretation, whereas Frisson and Pickering (1999) found no difficulty with interpreting Vietnam as an event. What is the difference between these cases, and what might it tell us about semantic composition?

Enriched composition appears to involve two operations: type-shifting itself and the insertion of new semantic structure. In started the book, the processor has to reinterpret the complement as an event and to insert the relevant meaning (e.g., reading). The experiments on metonymy (Frisson & Pickering, 1999; Pickering & Frisson, 2003) suggest that the act of type-shifting is not costly, so long as the type-shifted sense (e.g., Vietnam meaning the Vietnam War) is familiar. If so, it must have been the act of inserting the new semantic structure that was costly.

However, a direct comparison between the metonymy experiments and the coercion experiments is risky, because the former involved proper names, whereas the latter involved definite descriptions. To be more confident about the difference between metonymy and coercion, and hence to be able to conclude that coercion involves an on-line process, we needed to make a direct comparison between the processes. Hence, we considered items with involving a literal expression (10a), a metonymic expression (10b), or a metonymic expression (10c) in a coerced context (McElree, Pickering, & Frisson, 2003).

10a. The gentleman spotted Dickens while waiting for a friend to arrive.
10b. The gentleman read Dickens while waiting for a friend to arrive.
10c. The gentleman started Dickens while waiting for a friend to arrive.

If type-shifting per se is the cause of processing difficulty, then (10b) and (10c) should be harder to process than (10a). If, instead, difficulty is caused by the insertion of new semantic structure, then (10c) should be difficult, but (10b) and (10a) should not differ in difficulty. In fact the results supported this latter prediction. Largely on the total time measure, (10c) was harder than (10b) and (10a), which did not differ greatly in difficulty. These results therefore support the claim that familiar type-shifted senses can be stored, but that enriched composition has to take place on-line.

Conclusions

Our experiments suggest that type-shifting per se isn't costly, but that the on-line generation of non-lexicalized senses is difficult. In during Finland, people have to generate an event sense for Finland or else come up with no interpretation for the expression. Similar conclusions hold for reading Needham, in the absence of contextual support. Likewise, the generation of the appropriate sense for started the book causes difficulty. The fact that similar difficulty holds for started Dickens confirms our interpretation of the results.

Our experiments have generally found that these effects are "late" effects. In this respect, they contrast with frequency effects, some lexical effects (such as those that occur during the resolution of lexically ambiguous nouns), and some syntactic effects (such as the more extreme form of garden-path effects; e.g., Frazier & Rayner, 1982). The pattern of results also contrasts with answering to the stadium, where difficulty emerged immediately. It appears that some aspects of lexical, syntactic, and semantic processing do (largely) respect the immediacy and eye-mind assumptions (with some important caveats), but that many aspects of sentence interpretation are somewhat delayed. Pickering and Frisson's (2001) findings confirm this, in that it takes time to settle on an interpretation of a sense-ambiguous verb. It may be that people have to perform extensive searches for semantic information (e.g., what event might have occurred in Finland?), and that this takes time. It may be that different experiments produce delayed effects for different reasons, and we must be mindful of the possibility that the absence of some early effects might simply reflect a failure to detect a real effect (perhaps because of insufficient power). But the pattern of results suggests that not all "deep" aspects of semantic processing have effects that emerge during initial processing of the relevant word.

Which hypothesis should we therefore relax? It appears that the main problems are with the immediacy hypothesis, and that some aspects of processing simply take more time than the eye is "prepared to wait". Prima facie it makes more sense to relax this hypothesis than the eye-mind hypothesis, because the assumption that people look at the thing they are thinking about extends far beyond linguistic processing. But more importantly, most of the late effects discussed above are relatively well localized during second-pass processing. In other words, readers do

21

look at difficult words such as <u>Finland</u>, but they largely do this during later processing rather than during first pass.

The experiments discussed in this chapter provide some evidence that eye movements allow us to separate out different stages of processing. Because semantic processing may not all occur "at once," and because eye movements are sensitive to different stages in comprehension, eye-tracking experiments can help us determine what happens when and use this information to drive theoretical accounts of language comprehension.

Author notes

References

Brysbaert, M., & Mitchell, D.C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. Quarterly Journal of Experimental Psychology, 49A, 664-695.

Clark, H. H., & Haviland, S. (1977). Comprehension and the given-new contrast. In R. Freedle (Ed.), Discourse Production and Comprehension (pp. 1-40). Hillsdale, NJ: Erlbaum.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22, 1-39.

Copestake, A. and Briscoe, E. (1995). Semi-productive polysemy and sense extension. Journal of Semantics, 12, 15-67.

Duffy, S.A., Morris, R.K., & Rayner, K. (1988). The interaction of contextual constraints and parafoveal visual information in reading. Journal of Memory and Language, 27, 429-446.

Ehrlich, S. F., & Rayner, K. Contextual effects on word recognition and eye movements during reading. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 641-655.

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

Frazier, L. (1999). On sentence interpretation. Dordrecht: Kluwer.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. Journal of Memory and Language, 29, 181-200.

Frisson, S., & Pickering, M.J. (1999). The processing of metonymy: Evidence from eye-movements. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 1366-1383.

Frisson, S., & Pickering, M.J. (2001). Obtaining a figurative interpretation of a word: Support for underspecification. Metaphor & Symbol, 16, 149-171.

Jackendoff, R (1997). The architecture of the language faculty. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). Foundations of language. New York: Oxford University Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. Psychological Review, 87, 329-354.

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. Journal of Memory and Language, 45, 259-282.

Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: Conceptual relations of polysemous senses. Journal of Memory and Language, 47, 548-570.

Konieczny, L., Hemforth, B., Scheepers, C. & Strube, G. (1997). The role of lexical heads in parsing: Evidence from German. Language and Cognitive Processes, 12, 307-348.

Liversedge, S.P. (1994). Referential context, relative clauses, and syntactic parsing. Unpublished Ph.D. thesis, University of Dundee.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time In G. Underwood (Ed.), Eye Guidance in Reading and Scene Perception (pp. 55-75). Oxford: Elsevier.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. Nature, 244, 522-523.

McElree, B., Pickering, M.J., & Frisson, S. (2003). Enriched composition and sense extension. Manuscript in preparation.

McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched semantic composition. Cognition, 78, B15-B25.

Mitchell, D.C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), Attention and performance XII (pp. 601-618). Hillsdale, NJ: Erlbaum.

Pickering, M.J. (1999). Sentence comprehension. In S.C. Garrod, & M.J. Pickering (Eds.) Language processing (pp. 123-153). Hove: Psychology Press.

Pickering, M.J., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye-movements. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27, 556-573.

Pickering, M.J., & Frisson, S. (2003). Learning novel senses by rule: Why reading Needham can be easy. Manuscript submitted for publication..

Piñango, M. M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax-semantics interface. Journal of Psycholinguistic Research, 28, 395-414.

Pustejovsky, J. (1991). The syntax of event structure. Cognition, 41, 47-81.

Pustejovsky, J. (1995). The generative lexicon. Cambridge, MA: MIT Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372-422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition, 14, 191-201.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 779-790.

Rayner, K., & Pollatsek, A. (1989). The psychology of reading. Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line comprehension processes. Language and Cognitive Processes, 4, 21-49.

Rodd, J..M., Gaskell, M.G., & Marslen-Wilson, W.D. (2002). Making sense out of ambiguity: Semantic competition in lexical access. Journal of Memory and Language, 46, 245-266.

Sanford, A.J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence, Trends in Cognitive Sciences, 6, 382-386.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. Cognitive Psychology, 14, 489-537.

Sereno S.C. (1995) Resolution of lexical ambiguity: Evidence from an eye movement priming paradigm. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 582-595.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. Journal of Verbal Learning and Verbal Behavior, 18, 645-659.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. <u>Journal of Verbal Learning and Verbal Behavior, 18</u>, 427-440.

Traxler, M.J., Bybee, M.D., & Pickering, M.J. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. <u>Quarterly Journal of Experimental Psychology, 50A</u>, 481-497.

Traxler, M.J., McElree, B., Williams, R., & Pickering, M.J. (2003). <u>Context Effects in Coercion: Evidence from Eye-Movements</u>. Manuscript submitted for publication.

Traxler, M.J., Pickering, M.J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. <u>Journal of Memory and Language, 47</u>, 530–547.

Traxler, M.J., Pickering, M.J., & Clifton Jr., C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. <u>Journal of Memory and Language, 39</u>, 558-592.

Van Gompel, R.P.G., Pickering, M.J., & Traxler, M.J. (2001). Making and revising syntactic commitments: Evidence against current constraint-based and two-stage models. <u>Journal of Memory and Language, 45</u>, 283-307.