6 Comprehending written language:
a blueprint of the reader

Charles A. Perfetti

6.1 Introduction

Reading is a language process that begins as a visual process. From a psycholinguistic point of view, reading is a secondary language process, partly derivative of primary spoken language processes. Thus, to understand the cognitive processes of reading one must refer to three general processing phases: (i) the visual; (ii) the processes that convert the visual input into something else (a linguistic representation); and (iii) the processes that then operate on the encoded representation.

In examining what cognitive research has learned about these reading processes, this chapter emphasizes the second and third processes. The initial visual processes, critical as they are, constitute only an initial input to the reader, and do not in themselves constitute a language process. The processes that operate on the encoded representations have a complementary character. They are likely to be shared with the language processes that are examined in other chapters in this volume, for example syntax (see Cutler and Clifton, Chapter 5). To this extent, the processes that are unique to reading are those that carry out the transcoding of the visual input into linguistic (phonological) forms, which then make a range of meaning and grammatical features available to a language comprehension process. However, these linguistically-based comprehension processes are not the end of the story, because they may be supplemented by highly generalized (non-linguistic) processes in both written and spoken language comprehension.

This approach to written language has the virtue of simplicity, because, by its coupling to spoken language processes, it does not require a theory of comprehension that is unique to reading in all its aspects. However, there are two additional considerations that imply limitations on the effectiveness of written language comprehension, relative to spoken language comprehension; one is that the written language comprehension processes can take advantage of general language processes only to the extent that the visual-to-linguistic transcoding processes are effective. The second is
that the typical forms of written texts may place demands on comprehension that are not shared by spoken language. It is quite possible that additional comprehension strategies develop with literacy.

6.2 A blueprint of the reader

Following the lead of Levet (1989) in his analysis of speech production, Fig. 6.1 shows a general schematic view or ‘blueprint’ of the reader. The blueprint represents the information sources that a reader would be expected to use in gaining comprehension of written language. The use of this information must be considered from two perspectives—from the view of the writing system that provides the units of reading and from the view of the cognitive processes that result in reading.

6.2.1 Writing system factors

Consider first the writing system, which determines in a general way how written units connect with units of language. It is traditional to consider writing systems as falling into one of three categories—alphabetic, syllabic, and logographic. Alphabetic systems, including English, Italian, and Korean, encode their language at the unit of the phoneme. Thus, the smallest unit of the writing system, the letter, corresponds to a meaningless unit of speech, the phoneme. Syllabic systems encode their language at the level of the syllable. Thus Japanese Kana, the most well known of the syllabic systems, has basic writing units that correspond to spoken syllables. Logographies are traditionally considered to be represented by Chinese, on the assumption that the basic units of the writing system, the character, corresponds to a word in the spoken language. Although this designation is reasonable, it is a bit misleading, because the characters themselves are typically compounds whose components often correspond to meaning (semantic radicals) and pronunciations (phonetics). Furthermore, the character nearly always corresponds to a single syllable in spoken language. Thus it may be more accurate to refer to Chinese as a morpho-syllabic (DeFrancis 1989) or even morpho-phonological system.

The writing system is of considerable importance in an account of reading processes. Because the writing system provides units that map onto one or more levels of the language—phonemic, syllabic, morphological, lexical—it will influence the process of identifying written words. Within the broad constraints provided by the writing system, there are further influences provided by the details of the orthography—the system that actually implements the writing system. In the case of alphabetic systems, which are the most common, we can distinguish among Roman, Cyrillic, Korean, Arabic, Hebrew, Greek, and other alphabets. These variations produce very salient visual differences in the forms of the scripts. Of perhaps more importance than this variation in alphabetic form, however, is variation in orthographic structure. For example, the ‘Roman’ alphabet is used in writing English, French, German, and Italian, among other languages. But the orthographies in these languages differ in transparency or ‘depth’ (Frost et al. 1987). Italian is highly transparent (or ‘shallow’) in that the spellings correspond reliably to the phonemes of words; English is less transparent in that its spellings do not as reliably map onto its phonemes. Thus, although the spellings of chair and choir differ in only one letter, their pronunciations differ in all but the final phoneme. The change in English pronunciations with identical spellings, however, sometimes preserves morphology, as when national preserves the root spelling of nation while altering the first vowel sound.
Figure 6.1 represents the influence of writing systems and orthographies at an abstract level. To instantiate the system for the case of English, we would merely substitute ‘letters’ for the orthographic units, and the relevant connection would be from letters to phonemes. The variable mapping between letters and phonemes would not be directly represented. Instead, we would enrich the size of the units in both the orthography and the phonology, from single letters through digraphs (ch), letter strings (choir), and even whole words (choir). At some large enough unit size, the mapping between spelling and pronunciations tends to stabilize, leaving only a handful of indeterminate pronunciations—for example wind, lead. These ambiguities remain at the level of word identification—is it the word wind, as in ‘cold wind’ or the word wind as in ‘wind up’? One of the interesting characteristics of writing systems in general is that they produce relatively few of these ambiguities at the word level. In isolation, the relationship between written forms and spoken forms is more predictable than the relationships between forms and meanings, which tend to be more context dependent. Relatively strong form–form relationships may be one reason to expect phonology to be activated in reading in most writing systems (Perfetti and Zhang 1995a).

6.2.2 Elementary reading processes in a representation account

Given writing system factors, the process of reading can now be viewed from the perspective of the processes that convert the units of the writing system into mental representations, or understandings of printed messages. The comprehension of a printed message, must, to a considerable degree, depend on two major language-based components: (i) the identification of words and (ii) the engagement of language processing mechanisms that assemble these words into messages. The mechanisms provide contextually appropriate word meanings, parse word strings into constituents, and provide inferential integration of sentence information into more complete representations of extended text. As indicated in Fig. 6.1, these representations, to be clear, are not the result of exclusively linguistic processes, which must be complemented by a variety of general and specific knowledge sources.

It is visual word identification, however, that is the most distinctive process for reading. Elementary processes of word identification begin with a visual input, as shown in Fig. 6.1. The visual input—a string of letters—is handled by elementary perceptual processes that respond to various basic features (lines, angles, contours) and the relationships among them that define specific letters. The outcome of this process begins the reading process, the activation of grapheme units (individual and multiple letters), that constitute words. It is important to emphasize that the interdependence of the elementary processes is an empirical matter, and the blueprint shows them as a series of processes as a convenience.) In traditional models of human cognition, the words are represented in a lexicon, the reader’s mental representation of word forms and word meanings. Successful word reading occurs when visual input from a string of letters, activating one or more word forms in the lexicon, results in the word corresponding to the input letter string (rather than some other word) being identified. Along the way, phonological units, including individual phonemes associated with individual letters, may also be activated. The phonological units may increase the activation of word units, effectively bringing about their identification. It is common to refer to such a process as ‘phonological mediation’, on the assumption that the identification of a word has been mediated by a phonological recoding of the graphic input. It is also common to represent two pathways, one from graphemic units to meaning directly, and one from graphemic units to phonological units, and then to meaning (the mediation pathway).

However, Fig. 6.1 indicates these two possibilities in a different way, rather than as alternative pathways. The identification of a word involves the immediate co-activation of graphemic and phonological constituents. Semantic activation quickly begins as well and operates in a bidirectional mode from lexical representations (the lexicon) to the elements of the orthography and phonology that are immediately activated. In principle, it is possible for the graphemes to activate the word directly without an intervening stage of phonological mediation, and the degree to which phonological mediation occurs is an empirical question. Figure 6.1 represents a conclusion, based on one reading of the research, that phonological activation (orthographic–phonological coherence) will be immediate, even if, in some sense, it is not decisively instrumental in every instance of identification. The output of the identification process is a word form and some of its associated meaning(s) and other information (grammatical form, argument structure, thematic range) that might be needed for sentence comprehension. The comprehension of what is read is a process of assembling words that are identified into phrases and sentences—parsing—and building representations of text meanings. These processes take word identification as input and will be taken up for discussion in a later section, after fuller examination of the word identification component.

To emphasize one general point about the blueprint, the configuration of processes into boxes and arrows is not a commitment to a cognitive architecture, but rather a necessary and conventional manner of representing the sources of information that are used during reading. Generally speaking, the bi-directionality of arrows indicates the opportunities for important feedback information that goes from higher to lower representational levels. The constraints on feedback (its effectiveness and timeliness) in each case is an object of empirical investigation and there is no reason to assume that all cases are equal in their constraints. For example the feedback from discourse-level representations to syntactic representations is probably not as constraining as the feedback from phonological representations to orthographic representations. More generally, the blueprint allows for modularity of structures to the extent that some of the feedback loops are ineffective.

6.2.3 Non-representational (emergent) accounts of identification

The description so far centres on processes of activation, but refers also to representations of words in the lexicon. There is an alternative description of the events of
word identification in which words are not represented permanently in the mind of the reader but rather emerge from the patterns of activation. There are several alternative descriptions of these emergent non-representational systems, including parallel distributed (PDP) networks and recurrent-network resonance models. PDP networks model word identification through activation that is sent across layers of graphic input units, 'hidden' units, and phonological output units (Seidenberg and McClelland 1989; Plaut et al. 1996). Words emerge from the distributed patterns of activation.

Resonance models, based on a dynamic systems framework, represent word identification as the stabilization of dynamic patterns that are continuously modified by interactions among inputs and various dynamic states resulting from prior experience (Van Orden and Goldinger 1994). In a resonance model, the identification of a word emerges as patterns of activity in a recurrent network move towards stabilization. One of the interesting features of the Van Orden and Goldinger (1994) model is that patterns of graphic–phonological activation stabilize more rapidly than do patterns of graphic–semantic activation. In effect, a word form becomes identified primarily through the convergence of orthography and phonology. Meaning is slower to exert an influence on the process.

Emergent models of identification, whether based on dynamic systems or PDP frameworks, offer significant contrasts with symbolic accounts of identification. To some extent the differences among the models are less in their empirical correctness, which is difficult to establish, than in their heuristic values. Thus, emergent accounts expose interestingly different ways to conceptualize language processes, including reading. One of these differences is the importance of feedback mechanisms in word reading. Such mechanisms are strongly implied by results that show effects of phonology-to-spelling consistency in reading (Stone et al. 1997). For example, the phonological unit /ob/ is always spelled -ope at the end of a one-syllable word, e.g., probe. By contrast, the phonological unit /ip/ is spelled inconsistently, sometimes -eap as in heap, sometimes -eep as in deep. Thus, Stone et al. (1997) found that a word such as probe is read more quickly than a word such as heap in a lexical decision task. Such results indicate that word reading models, whether their architecture is symbolic or emergent, need to include feedback from phonology to orthography as well as feedforward (orthography to phonology) connections. Models that describe emergent identification processes may seem especially compatible with the development of neurocognitive models of reading. Their ability to capture systems in states of change, whether through recurrent networks or layered networks of distributed information, may lend themselves to neurological modelling in a fairly natural way.

Nevertheless, there appears to be little, so far, that would favour one class of models over another just on cognitive grounds. Modelling systems are largely underdetermined by behavioural data, and the models are subject to fine tuning that can increase their ability to handle otherwise challenging results. However, some of the differences between symbolic and emergent accounts of word reading highlight important empirical questions about word identification, as illustrated in the next section.

### 6.3 Issues in word identification

The foregoing provides general descriptions of word identification, with variation in descriptions being a matter of architectural frameworks. Needless to say, there are some interesting questions about the details of word identification processes that lie beyond this general description, and in some cases beyond the general consensus that underlies it. Some of these questions are highlighted by the alternative approaches—symbolic and emergent—described above. Others amount to important details that must be addressed by any system.

#### 6.3.1 Routes to the lexicon

An enduring question, and one with obvious neurocognitive implications, has to do with whether there are two distinct pathways to the identification of words. Dual Route Theory embodies the hypothesis that there are indeed two routes: one route provides direct contact to a word representation from the graphic input. The second route converts graphemes, singly or in strings, into phonemes, which are used to access the word representations. The first route is sometimes referred to as the 'addressed' route, and the second, as the 'assembled' route. The addressed route is the direct look-up of an address; the assembled route is the assembly of the word's phonology as an intermediate step in the process.

Dual Route Theory, which was advanced by Coltheart et al. (1977) and elaborated in subsequent research (Besner 1990; Paap and Noel 1991), takes the following problem as central in word identification and solves the problem with the dual mechanism model: some words contain predictable spelling–sound correspondences, and some do not. Those that do, such as safe, can be read by converting the letters to phonemes, assembling the phonemes, and matching the assembled phonology with a lexical entry. But safe can also be read by an addressed mechanism, which simply looks up the lexical address of the letter string s-a-f-e. However, for the word cafe, the assembly process fails; by following the grapheme–phoneme correspondence rules for English, the assembly mechanism produces [kef] instead of [kəfe], an error. The only way to get the word right is to look up the spelling and produce the correct pronunciation. By itself this fact would not require two routes. A direct route could, in principle, be used for both safe and cafe. What seemed to force the need for the assembled route is the ability to pronounce non-words, such as zate or tafe. Because there is no word to look up, a process of assembly appeared to be necessary.

Additional data provided pillars for Dual Route Theory. One is the fact that in research on naming times for printed words, word frequency and word regularity interact; regularity affects the naming of low frequency words more than high frequency words (e.g. Seidenberg et al. 1984). Dual Route Theory handles this fact readily with the addition of two assumptions: that both pathways are activated in parallel and that the direct pathway is faster than the assembled pathway. Thus, for high frequency words, the direct pathway usually wins, even for regular words. For low frequency words, the direct route is a bit too slow, giving the assembled route a chance to compute
the phonological form first. An irregular word must lag behind to wait for the direct route, because its irregularity creates noisy output along its assembled pathway. This general result has been replicated many times and has been studied by clever manipulations designed to modify the effectiveness of the two routes (Paap and Noel 1991). Additionally, patient data showing relatively selective impairment of either non-word reading (with high frequency word reading intact) or of irregular word reading (with non-word reading intact) have been argued to require the Dual Route Model (Coltheart et al. 1993). Brain lesions in such cases appear to have disturbed either only the assembled pathway or only the addressed pathway rather than produced a general impairment in word identification.

Although Dual Route Models handle such phenomena in a straightforward way, it turns out that single mechanism emergent models are also successful. In the case of word identification, the PDP models, armed with hidden layers and back-propagation algorithms that use feedback about correct pronunciations, can actually produce the major phenomena. In fact, the feedback mechanism allows the models to learn from exposure to printed words, which, with correct feedback, alters the weights assigned to graphic input–phonological output pairs (through the computational hidden units). Thus, as exposures to word forms come to reflect the actual frequency of their occurrence, the model learns to pronounce them correctly most of the time. Moreover, novel words and non-words are pronounced by using the same mechanism. This is achieved, in effect, by initially following whatever leads are provided by the network’s learned connections, but then modifying the weight for a given form from its own feedback. In a series of simulations, Seidenberg and McClelland (1989) demonstrated impressive performance by a PDP network built from 400 graphic input units and 460 phonological units connected through 200 hidden units. It produced, in particular, the important frequency × regularity interaction. In effect, the phonological feedback for consistently pronounced letter patterns allowed ‘regularity’ to emerge with learning; and the high frequency established by specific word inputs was powerful enough to negate these regularity effects for high frequency words.

The only serious shortcoming of the single mechanism model was its relatively poor ability at reading pseudowords, which was demonstrated by Besner et al. (1990). Whereas human readers could read pseudowords accurately (around 80 per cent), the model could not. Seidenberg and McClelland (1990) argued that this failure reflected only the fact that the model had not been exposed to the number of words that a skilled reader would have been. With more exposure, pseudoword reading would have been better. More recent examples of PDP modelling have reported more success in pseudoword reading, an ability to model selective impairments, and additional performances that further narrow the gap between Dual Route Model and PDP successes (Plaut et al. 1996). (See Coltheart et al. (1993) for a detailed argument for the advantage of Dual Route Models over the earlier PDP models.)

At this point, an objective assessment of the Dual Route Model is that it continues to provide a good account of critical data in the field. But the PDP models are able to do quite well with a single mechanism and no explicit lexicon. The PDP models are powerful enough, in principle, to adapt their learning to a variety of data. The prospects for ‘critical’ behavioural experiments that can distinguish between single mechanism and dual mechanism models may be diminishing.

6.3.2 Phonological mediation

Notice that in the Dual Route Model, phonology occurs only along the assembled route. Thus phonology appears to be an optional component of word identification, taken only when the assembled route is used. However, even when the addressed pathway is the one that brings about word identification, some assembly of phonology can occur. It just happens to be the ‘wrong’ phonology for word identification. In the single mechanism PDP models, phonology can be thought of as continuously occurring, as connections between graphic input units and phonological units (through a hidden layer) are activated.

Nevertheless, the role of phonology has remained controversial in word identification. Is it optional? Is it obligatory? Is it causal, mediating identification? Or ‘post-lexical’, resulting from identification? The results of research have been mixed and open to a variety of interpretations. As a general rule, tasks that limit the exposure of a letter string and mask its presentation may show phonology more reliably than tasks that do not (Berent and Perfetti 1995). For example, in backward masking paradigms, a word is presented briefly, 20–50 ms, and followed by a letter string mask. When the following mask is a pseudoword that shares the phonological form of the target, as in rake masked by rak, the identification of the target, which is adversely affected by the mask, is more likely than when the word is followed by a graphic mask, for example ralk, that shares the same number of letters with the target but not as many of the phonemes. The graphic mask produces superior identification of the target compared with a control mask that does not share letters with the target (Perfetti et al. 1988; Berent and Perfetti 1995). Thus there are two effects, each suggesting that the identification of a word includes sublexical constituents: (i) abstract graphic units (because letters shared between target and mask aid identification even when they are in a different font) and (ii) phonemic units (because phonemes shared between target and mask additionally aid identification). Similar effects are obtained when the presentation of the pseudoword precedes the target in a priming procedure (Perfetti and Bell 1991). (Priming in brief exposure situations can also be considered to be forward masking in that the ‘prime’ masks the following target word.) Because these brief exposure paradigms, both backward masking and priming, actually interrupt word processing, they can expose the partial products of the identification process—the sublexical units that are activated on the way to identification. Perfetti and Bell (1991) interpreted their effects as demonstrating phonemic processing within the first 40 ms of exposure to a word. Similar experiments in ‘form priming’ of lexical decisions and word naming produce evidence concerning the effects of graphic forms (Forster and Davis 1984, 1991). There is now considerable evidence from brief exposure paradigms that supports the conclusion that phonological as well as graphemic units are activated in the
reading of alphabetic systems (Ferrand and Grainger 1992; Grainger and Ferrand 1994; Lukatela and Turvey 1990a,b).

Interestingly, the results from these brief exposure paradigms not only implicate sublexical orthography and phonology, they do so for high frequency as well as low frequency words and across variations in spelling-to-phonology consistency. This contrasts with the selective effect of regularity seen in word naming experiments. This difference between full exposure and brief exposure paradigms in their sensitivity to sublexical phonology was reviewed by Berent and Perfetti (1995). In explaining the differences between full and interrupted exposure to words, Berent and Perfetti proposed the Two Cycles Model of word identification, in which consonants and vowels are assembled from words in temporally distinct processing cycles. Although the linear string of letters may undergo serial left-to-right processing at the graphic input stage, the model assumes that the phonology is assembled non-linearly. Consonants (i.e. phonological consonants) are assembled in a first cycle and vowels (i.e. phonological vowels) are assembled in a second cycle. In effect, the model claims that consonants, which are more regular than vowels, always are assembled 'prelexically', that is prior to lexical access. Frequency, a lexical effect, and regularity effects arise at the second vowel cycle. Experiments that vary the stimulus-onset asynchrony (SOA) between target and mask, as well as the specific source of target–mask overlap (consonants vs. vowels), provide some support for this assumption.

It is not the case, however, that only brief exposure paradigms provide evidence for generalized sublexical phonology. Van Orden (1987) found a phonological interference effect when subjects were required to make category judgements. For example, when rows was presented as a foil for the category flower, decision times were longer and prone to errors. However, there may be some situations that do not produce this effect (Jared and Seidenberg 1991). Among other non-naming tasks that have provided evidence for automatic phonology in word identification are letter detection (Ziegler et al. 1997) and lexical decision (Stone et al. 1997). One of the most compelling classes of evidence comes from experiments with Serbo-Croatian, which is written in two different shallow orthographies, one using the Roman and the other, the Cyrillic alphabet. These experiments, across a number of paradigms, produce results that can only be explained by the assumption that grapheme–phoneme connections are activated at every opportunity during simple word reading (Lukatela and Turvey, 1998).

The question of mediation implies more than phonological activity. It has referred traditionally to a causal relationship between the activation of a phonological form and some other identification event, especially the access of word meaning. It is one thing to discover sublexical phonology in word identification; it is another to demonstrate the instrumentality of this phonology. For this, Lukatela and Turvey (1991) and Lesch and Pollatsek (1993) developed semantic priming paradigms that expose phonological mediation. Semantic priming has been well demonstrated in naming, so, for example, presentation of beach would facilitate the naming time for the following target word sand. Phonologically mediated priming occurs when the prime is not beach but its homophone beech. Beech primes sand through the activation of its phonology, which in turn activates beach, and primes sand through its semantic link.

Lesch and Pollatsek discovered that when the prime was exposed for 50 ms, followed by a pattern mask of 200 ms, the priming by the homophone was as large as the priming by the semantic associate, which they interpreted as demonstrating mediation—meaning access produced by phonology. But when the prime was presented for 200 ms, followed by a mask of 50 ms, mediated priming disappears as a spelling verification process occurs (Van Orden 1987).

There is an alternative to the traditional view of mediation as a causal event in access to word meaning. Rather than an instrument of access, phonology can be considered to be a constituent of identification (Perfetti and Zhang 1995b). The assumption is that the phonology always co-occurs with a word’s graphic and meaning constituents, to constitute a three-constituent word identity. In a resonance model framework, which does not partition perception into discrete events such as implied by traditional mediation, Van Orden and Goldinger (1994) argue that mediation can be viewed as a process of stabilization in which word identity is ‘negotiated’ between sources of word form information. This rapid stabilization of word identity is enabled by the fact that orthography is more reliably mapped to phonology than it is to meaning.

6.3.3 The time course of graphic, phonological, and semantic activation

The mediation issue, especially when viewed as the stabilizing effect of rapid phonology on word identity, raises the question of the time course of activation of word constituents. The graphic, phonological, and meaning information sources that come together in the identification of a printed word become available over a very brief time period. The time to identify a printed word, as assessed in eye-tracking research, brings estimates of 200–300 ms (Rayner and Pollatsek 1989). On the other hand, in brief exposure paradigms with masked presentation to disrupt identification, 50 per cent thresholds are reached within 50 ms or so. Although such data may suggest that the visual system does not need very much time to initiate the process of word identification, visual transmission across areas of visual cortex requires more time than this, based on single cell recordings in non-human primates (Nowak et al. 1995). Thus, processing time estimates obtained in behavioural paradigms must be interpreted not as absolute neuroprocessing times, but as estimates of how quickly various information sources become available relative to others in specific processing tasks.

The question of when graphic, phonological, and semantic information becomes available is a difficult one to answer, dependent both on tasks and models. For example, the resonance model of Van Orden and Goldinger (1994) requires that phonological information immediately coheres with graphic information; there is no time at which the system has only graphic information. In a standard model, however, the process, which of course must begin with a graphic input, can take several courses. Partial graphemic information can activate both phonological and semantic information.
associated with candidate words, which can in turn feedback to graphic information. Interactions among levels (Plaut et al. 1996), rather than simple feedforward of information, suggest that, generally, there should be overlapping activation cycles of semantic and phonological information within a fully interactive system. However, the timing of specific events will be very sensitive to a wide range of factors: the printed frequency of a word, the spoken frequency of a word, its orthographic and phonological length, the consistency of its spelling, and its range of meaning possibilities. For example, a specific word meaning might be activated more quickly than phonological information for a word with high printed frequency, an inconsistent spelling pattern, and a single well-defined meaning. Phonological information should be more rapid for a word with low printed frequency, a consistent spelling pattern, and more varied meaning.

Research has identified the relative time course of orthographic and phonological word constituents in both brief exposure and full exposure paradigms. With lexical decision tasks, Ferrand and Grainger (1992, 1993) found orthographic facilitation with as little as a 17 ms prime exposure, with phonological facilitation emerging at around 50 ms. Perfetti and Bell (1991) found that subjects could use phonological information shared between a prime and target within 45 ms of prime exposure, only 10 ms of time lag relative to the use of shared graphemic information. No time lag between orthographic and phonological benefits was found in backward masking, with both benefits found within 35 ms. When words rather than pseudowords are used as primes, both facilitation and inhibition (reduced target identification) are produced. Tan and Perfetti (in press), examining the time course question in the backward masking paradigm with real word masks, found the earliest facilitative effect for graphic information coincided with phonological inhibition when both target and mask were presented for 28 ms. When the target was exposed for 42 ms, followed by a mask of 28 ms, both graphemic and phonological facilitation were obtained. Associative masks, on the other hand, began to inhibit target identification. Generally, with only slight variation across tasks, the picture seems to be that phonological information is rapidly available in word processing, only slightly later than orthographic information.

In non-alphabetic writing systems, studies of the time course of Chinese word components have not found evidence of semantic information prior to phonological information. Perfetti and Zhang (1995b) found that meaning judgements about pairs of Chinese characters produced phonological interference within 90 ms; judgements about pronunciation produced semantic interference within 140 ms. In primed naming tasks, where the time course is assessed according to various prime-target relations (graphic, phonological, semantic) graphic information is the first available, followed by phonological and then semantic (Perfetti and Tan 1998). Thus, whether in naming and or in non-naming judgements, the potential of the Chinese writing system to allow the by-pass of phonology in a direct-to-meaning process, does not prevent readers from encoding the phonology of printed words. In fact, the evidence from time course studies so far suggests it is semantics, rather than phonology, that is delayed at the character level. The stronger bonding of graphic and phonological forms, compared with graphic forms and meaning, may help explain this (Van Orden et al. 1990; Van Orden and Goldinger 1994; Perfetti and Zhang 1995a,b). Even Chinese may have more reliable mapping from visual form to phonological form (at the character level) than from visual form to meaning.

The time course of processing orthographic and phonological constituents of words is also a question amenable to temporally sensitive neurocognitive methods, specifically the recording of changes in event-related electrical potentials from the surface of the scalp. As discussed elsewhere in this volume, linguistic processes have been associated with specific ERP components, including the N400, a negative potential with an onset at about 200 ms after stimulus onset, which has been interpreted as an indicator of semantic analysis or of the integration required by prior context (Brown and Hagoort 1993). The more constraint the context puts on the target word, that is the more predictable the word is from the context, the lower the N400 amplitude. (For a review, see Kutas and Van Petten 1994.)

However, the N400 is sensitive not only to semantic events, but can be modulated by orthographic and phonological information and by the demands that tasks place on the use of this information. For example in rhyme judgements, the N400 is affected by both phonemic and orthographic similarity of the rhyming words; but in simple visual similarity judgements about word pairs, the N400 is affected only by orthographic similarity (Polich et al. 1983). Early effects of orthographic form are also seen when subjects make lexical decisions, with an early modulation of the N400 produced by orthographic overlap between successively presented primes and targets, even when the prime is a pseudoword (Doyle et al. 1996). As for phonological information, modulation of the N400 also has been reported during rhyme judgements to printed words, even when there is no orthographic overlap (e.g. shoe-cheer) (Rugg and Barrett 1987). Moreover, components other than the N400, both later and earlier, may be sensitive to phonological information. For example, one study of silent reading for meaning found that an earlier negative component (peaking around 200 ms after word onset) is associated with homophones (boat) of a word (bore) that would have been sensible in that context, but not with words orthographically (boat) related (Niznikiewicz and Squires 1996). Thus, ERP results show sensitivity to the timing of orthographic and phonological processing, and do not appear to contradict a rapid phonology hypothesis. However, tracing the relative time course of events during word identification requires careful attention to tasks as well as to the separation of the orthographic, phonological, and semantic properties of words.

In summary, although the time course question is complex and presumably task dependent, experimental results are consistent with the rapid phonology hypothesis. Although one might expect semantic information to become available as quickly as phonological information, there is little evidence for this in the research. Such a result is surprising only if one assumes that there is a visual-to-meaning process that doesn't involve the primary linguistic (speech) system. On the view that primary speech processes remain functional in reading as well as in spoken language, there is no reason to assume that phonology is either by-passed or delayed in identifying words.
6.3.4 Word meanings and word forms
So far, I have treated word meaning as an unanalysed concept. However, multi-morphemic words, whose meanings reflect semantic composition from more than one morpheme, are very common. Dislike, for example, is related to like and discredit, and undo is related to do and untie. Certainly readers, like speakers and listeners, have implicit knowledge of these morphological relations. For reading, the question is whether some kind of morphological decomposition process accompanies printed word identification. One view is that words are represented as full forms without references to their morphological constituents (Butterworth 1983; Osgood and Hoosain 1974). An alternative view, more widely held, is that morphemes contribute to word reading. Whether words are decomposed into morphological components before or after word recognition is a further question (e.g. Fowler et al. 1985; Feldman 1994; Taft and Forster 1975; Taft 1992). Whether the morpheme is a unit of processing and mental organization is the question, and this question has proved difficult to answer in a simple manner. However, it appears that readers can be quite sensitive to the morphological structure of words under some circumstances.

6.4 Reading words in context
6.4.1 Word meaning activation and selection
Semantic information becomes available as words are read. However, words have many meanings, and the selection of a functional meaning depends on context. For example in a sentence such as The men decided to wait by the bank, the word bank is ambiguous. The question is how a reader selects the relevant meaning—the intended meaning—in any given sentence. Of course the general answer to the question is context. The meaning of a word in a particular instance is determined by the context in which it occurs. In the above example, where the context does not appear to constrain the meaning that bank can take, the process of meaning selection would appear to be indeterminate. The reader might select either the meaning ‘financial institution’ or ‘side of a river’ or both. We might expect that lacking a helpful context, the selection of word meaning will depend on statistical structures: the reader will tend to select the meaning that is the more common, in this case bank as ‘financial institution’, or at least the one more common in the experience of the individual (compare bankers with fishermen).

In more constraining contexts the situation becomes more interesting. Suppose one of the sentences below is encountered:

1. I pulled the fish up onto the bank.
2. I opened a checking account at the bank.

In (1) the riverside sense of bank becomes more likely than its financial sense, and, inversely for (2). The question is now whether the selection of the relevant meaning of bank occurs without notice of the irrelevant meaning. One possibility is that the selection of meaning is so strongly determined by the context—or what can be called the message level—that the word’s meanings are selectively accessed; the one consistent with context is accessed and the one inconsistent with context is not (Glucksberg et al. 1986). This is the selective access model.

A second possibility, the multiple access model, is that the message level—the context—can assert no influence on the word level at first. On this account, the meaning selection process, which is determined by context, is preceded by a very brief general activation process, in which more than one meaning of a word is activated. This meaning activation process is automatic and very rapid (less than 250 ms), followed by a process that selects one of the meanings on the basis of consistency with message level information (Seidenberg et al. 1982; Kintsch and Mross 1985; Onifer and Swinney 1981). Thus, this account of autonomous activation + context selection is that both meanings of bank are initially activated regardless of the context, which then quickly selects from the activated meanings the one that fits the context.

A third possibility rests on limited multiple access that is controlled by the relative frequency of the word’s meanings. The ordered search hypothesis assumes the most frequently used meaning of a word is always activated (Hogaboam and Perfetti 1975; Duffy et al. 1988; Neill et al. 1988). Reordered search adds the assumption that context can exert a short-lived re-ordering of the word’s meanings (Duffy et al. 1988). Thus, on these accounts, the financial meaning of bank would be activated in both sentences (1) and (2), although it would be selected only in sentence (2). In an ordered search model, the contextual help provided for the less frequent river-bank meaning in (1) would not be sufficient to eliminate activation of the irrelevant dominant financial meaning. Refinements of the ordered search model are required to account for the fact that the relative frequency difference (meaning dominance) is a factor, as one would expect on the basis of a graded frequency (as opposed to all-or-none) assumption (Duffy et al. 1988).

The research surrounding these models has used a variety of procedures that vary in their potential for exposing the fine-grain temporal issues involved. Notice that the issue is never whether context exerts an influence, only when it does so. Lexical decision with cross-modal priming, in which the SOA between prime and target is varied, has been the most common method. One looks for whether a lexical decision target related to the unintended meaning of the word can be primed at very short SOA by the word in context. Eye-movement studies also provide a window on these issues by observing eye fixations on ambiguous words under various contextual conditions.

The research on these questions has grown quite large and resists simple summary conclusions. It does seem fair to say that whereas early research tended to support some version of multiple access (either ordered search or parallel activation versions), more recent research points to the possibility that context can exert a stronger and earlier influence under some situations (Kellas et al. 1991; Simpson and Kueper 1991; Tagiossi 1988, 1991). Important in these results is the emphasis on properties of the context that can or cannot bring about sufficient constraint to allow selective access (i.e. no activation of the alternative meaning) to even a less frequent meaning.
Nevertheless, the overall pattern of results in this field cannot be taken to support a general process of prior meaning selection by context. Instead, as Simpson (1994) concluded in a review of the research, the overall pattern of results can be explained by assuming that all meanings are activated, but with the degree of activation sensitive to both the relative frequency of the meanings and the context.

6.5 Understanding sentences

To move from the word to the text can be a very large step. However, understanding sentences requires the identification of words. Eye movement studies estimate that over 80 per cent of the content words (as opposed to ‘function’ or grammatical words) in a text are fixated when a reader reads for comprehension (Carpenter and Just 1983). Although the length of a word plays a large role in the probability that it will be fixated, the general point is that substantial word reading is required to understand written texts. Obviously, word reading is only the beginning of the process. In insisting on this truism, we must be mindful that it does not imply that comprehension proceeds simply and unidirectionally from word identification to comprehension. The preceding section makes clear the possibility that what is obtained from a word—its meaning—at some point in the process is determined by the message context. Because word identification, now taken to include the semantic constituents of a word, is influenced at some point by context (again the issue is not whether but when), there must be some feedback between processes that represent messages and processes that select meaning for words.²

To represent the broad picture of reading comprehension, Fig. 6.1 shows the general sources of information that, combined with word identification, lead to comprehension. In general terms, comprehension occurs as the reader builds a mental representation of a text message. In this building process, lexical, syntactic, and inferential processes contribute, all in some degree of interaction with non-linguistic knowledge. The architectural question in reading comprehension research has been whether these interactions take place in a system that constrains their influence.

One approach to this architecture question has been to assume that the processes that interpret sentences to yield messages are constrained either by specific properties of processing components or by constraints in the direction of information flow among the components. Thus, the modularity thesis (Fodor 1983) is that there is a set of specialized processors, including language processors, that have characteristics that resist influence from other processing mechanisms. A syntactic processor would have access to restricted information about the forms of syntactic constituents—noun phrases, verb phrases, etc.—and would go about the business of building syntactic representations based on its specific, restricted sources of information. Its output—some form of constituent structure—would be constructed without interference from general knowledge that might be helpful in determining what the sentence is about. In effect, modularity prevents expectations about messages to influence the construction of syntactic forms. A similar analysis results from assuming that the processors, whether or not they correspond to modules, carry out their work in a feedforward fashion, in which lexical outputs provide input into syntactic processors, which, in turn provide input into semantic processors. This analysis leads to the hypothesis of an autonomous syntax as well as an autonomous lexicon (Forster 1979). Thus, one can expect autonomous processors within message construction, either as a consequence of essential characteristics of cognitive system modules or as a result of the organization among the processors. In either case, the important point for the blueprint of the reader is that there is little or no opportunity for general knowledge (expectations, discourse context, etc.) to have an influence on the operation of the syntactic processor (the parser).

The alternative architectural view is that interactions among sources of information are quite unconstrained (MacDonald et al. 1994) or, at least, that some non-syntactic information sources have an early influence on sentence parsing (Crain and Steedman 1985; Altmann and Steedman 1988). Thus, the general empirical question on which the interactive and the autonomous syntax hypotheses differ is whether the initial parsing decisions—understood roughly as a momentary attachment of the words (or morphemes) onto a syntactic structure—are independent of context. Although there are a number of both general and detailed proposals for how the parser works, the minimal attachment principle of Frazier (1979); Frazier and Rayner (1982) illustrates the architectural issue clearly, and has been a focus of empirical work. The principle assumes that a reader, like a listener, attaches every word as quickly as possible to a syntactic configuration that is built one-word at a time through sentence processing. There is no postponement of a syntactic decision. Nor are alternative structures built. Given this assumption, the minimal attachment principle is that, at any point in processing a sentence, the word (or morpheme) is attached to an accumulated syntactic configuration (or syntactic tree) so as to create the simplest attachment possible, consistent with syntactic constraints. Simplicity is a matter of the number of nodes in a syntactic tree that have to be established.

To illustrate, consider the sentences below from a study by Ferreira and Clifton (1986):

(3) The defendant examined by the lawyer turned out to be unreliable.
(4) The defendant who was examined by the lawyer turned out to be unreliable.

In both (3) and (4), the reader must interpret ‘turned out’ as a main clause verb phrase, that is the defendant turned out… However, in (3) the reader, in accord with a minimal attachment principle, takes ‘examined’ to be the main clause verb, that is the defendant examined [something]… This incorrect initial decision then leads to a problem when ‘turned out’ is encountered, because there is no open verb phrase slot for it. Thus, (3) is a garden-path sentence that produces a processing slow down in the critical (disambiguating) region turned out compared with the same region in (4) (Ferreira and Clifton 1986; MacDonald et al. 1992). In (4) the occurrence of that, a relative pronoun, forces the parser to build the right attachments. In (3), the absence of that allows the parser to build either a main clause verb or a (reduced) relative clause when it
encounters examined. The main clause attachment is syntactically simpler, because it requires no additional syntactic nodes to be established. The relative clause structure is more complex, because it requires modifier to be established at the initial noun phrase—something like the defendant [ the defendant examined by the lawyer], where the bracketed material is an additional node in the syntactic tree.

One can readily imagine other proposals for why a reader might have trouble with (3) and not (4), and some of these refer to the assignment of thematic roles (Pritchett 1988, 1992). These roles are filled by nouns in relation to the verbs. Thus, examined needs to assign roles to an examiner (an agent) and, in the verb phrase, to an examinee (a theme). Sentence (3) allows the first noun to be chosen as subject/agent, and leaves additional roles to be filled after the verb. There is evidence that the thematic roles defined by the verb play a part in parsing decisions (Frazier and Rayner 1982; Britt 1994; MacDonald et al. 1994). These and related syntactic issues are discussed elsewhere in this volume (Chapter 5, Cutler and Clifton; Chapter 9, Hagoort et al.). Here, it may be enough to conclude that the field is far from having established a clear consensus on exactly how all the sources of information that might be used in parsing actually do get used.

To return briefly to the central empirical question, however, it is whether the problems with a sentence such as (3) can be avoided in the right kinds of contexts. Can the processing difficulty be eliminated by a context that encourages the reader to immediately take the first noun, the defendant, as the object of examined rather than as its subject? Avoiding the incorrect assignment in this particular structure (the reduced relative clause) is very difficult, and a number of studies have concluded this garden path reflects a very deeply embodied syntactically-based preference, whether minimal attachment or some other syntactic principle (Britt et al. 1992; Ferreira and Clifton 1986). However, some studies suggest the contrary, that the garden path—the momentary mistake in parsing that makes the defendant the subject rather than the object of examined—can be avoided by the use of some other information source: by properties of the initial noun itself—its meaning characteristics; by statistical properties of the verb and its past participle (examined); and by the referential context provided by a previous discourse segment (e.g. MacDonald et al. 1982; Trueswell et al. 1994). Additionally, some results suggest that susceptibility to the garden path is a function of the reader’s basic working-memory capacity, such that readers with large capacities can keep more than one parsing possibility in mind and then choose the needed one when later sentence information is read (Just and Carpenter 1992).

An example of how thematic roles and context interact is provided by Britt (1994), who varied the context and the verb type in post-nominal preposition phrases such as (5) and (6):

(5) He dropped the book on the war onto the chair.
(6) He put the book on the war onto the chair.

An important point about these sentences is that they produce garden paths that are much less severe than the reduced relative clauses illustrated in (3) and (4).

Both (5) and (6) tend to produce garden paths, in that readers’ initial decisions lead to attaching the prepositional phrase on the war as part of the verb phrase, whereas it needs to be attached as part of the noun phrase. Britt found that this garden path could be overridden by a context that favours the noun phrase attachment; such a context specifies two books and constrains the situation so that only one of them—the one on the war—will be dropped. However, this context effect works only for verbs such as drop (5). Such a verb requires only one argument to be filled following the verb. This argument is filled by the theme role, the object that is dropped. Thus one must specify a dropped object in using drop, but need not specify the end state of the dropped object. (Notice that John dropped is incomplete but John dropped the book is not.) However, for verbs like put (6), two arguments must be filled following the verb: an object (theme) that gets put, and a location that specifies its end state. (Notice that John put the book and John put on the floor are both incomplete or ungrammatical.) The very same favourable context that avoids the garden path with drop fails to avoid it with put. Thus Britt’s experiments suggest that the internal structure of a verb (its required arguments) is a more powerful determiner of parsing decisions in these structures than is the biasing effect of a discourse context.

The general question of how these various factors combine to determine parsing decisions remains very difficult. Again, one approach is to assume that parsing principles, derived from basic syntactic knowledge, control the first stages of comprehension and then are quickly (and usually beyond awareness) modified by other factors. An alternative account allows immediate use of all sorts of information, with no special processing provided by a syntactic analyser. For example the general proposal of MacDonald et al. (1994) is that a reader’s parsing decisions are determined by the history of encounters with various forms and functions. All of these frequency-based considerations and more are assumed to be important: the semantic properties of a word, the relative frequency of a verb in a particular grammatical role (e.g. examined as a past tense verb vs. as a past participle), the relative frequency of specific nouns as fillers of thematic roles associated with verbs (e.g. defendants as an agent of examine relative to a theme (object of examination)), and the actual frequency of specific syntactic structures (main clauses with versus without reduced relative clauses). A reader is more or less likely to be garden pathed as a function of these frequency-based characteristics. To the extent that these characteristics support a particular parsing decision, that decision is more likely than some alternative.

The fundamental principles that guide parsing remain a source of disagreement. Empirically, distinguishing among these principles turns on the timing of reading events, because the question is not whether context is used, but rather when it becomes available to the parser. While this question continues to be the object of research in studies that use self-paced reading and eye-tracking methods, which can provide reasonably good data on the time duration of reading events, temporally sensitive neurocognitive methods are also proving to be informative. Studies of event-related potentials (Hagoort et al. 1993; Neville et al. 1991; Osterhout et al. 1994) have discovered some distinct ERP components that are not associated with the N400 that has
been linked to semantic processing. For example, Hagoort et al. (1993) found a positive component associated with syntactic violations of certain kinds, a component they called the 'Syntactic Positive Shift'. Osterhout et al. (1994) found a similar component associated with ambiguities that would lead to garden paths when they are processed by the minimal attachment principle. The convergence of results from several different studies may point to an ERP component sensitive specifically to processes that determine constituent structures.

Finally, it is interesting to note that most of these parsing results have been obtained with visual situations, and so apply most directly to reading. However, the reader has information in conventional print not available to the listener (and vice versa). Commas and full points, in particular, tend to disambiguate structures, and their role in helping the reader avoid parsing problems has been examined in a few studies (Adams et al. 1991; Mitchell and Holmes 1985). Interestingly, the use of punctuation is not sufficient, by itself, to avoid at least some of the syntactic preferences that lead to garden paths.

6.6 Comprehension beyond sentences

The reader uses sentences to build an understanding of a text, and for that larger purpose sentence comprehension is only part of the picture. The reader must combine the message of each sentence with the message accumulated up to that point on the basis of the prior text. This appears to complicate the comprehension process only slightly in requiring (i) individual sentence comprehension and (ii) integration across sentences. However, behind each of these is a complex picture that must accommodate a range of processes that assign discourse referents (the things referred to in the text) to the elements of sentences, and that establish higher-level representations that capture the gist of the text as a whole. These text-level processes must appeal significantly to sources of information that lie beyond the verbatim representation of the text—the reader's knowledge about the semantic domain of the text, the type of text (genre)—and to inferential processes that assist the integration of information.

6.6.1 Mental representations of text

The skilled reader has not one but many levels of representation for a text. The two most general are a model of what the text says (the text base) and a model of what the text is about (the situation model). In the influential framework of Van Dijk and Kintsch (1983; Kintsch 1988), the text base is a mental representation of the propositions of the text. The atoms of meaning are extracted from sentences, built up through the reading of the successive sentences of the text and supplemented only by inferences necessary to make the text coherent. The reader builds a situation model from the text base by combining knowledge sources through additional inference processes. Thus, a text base is essentially linguistic, consisting of propositions derived from sentences, whereas a situation model is essentially agnostic in its form of representation. It may well be linguistic—an elaborated set of propositions that includes inferences as well as propositions extracted from actual text sentences. However, it may also be fundamentally non-linguistic, a mental model that directly represents referential and spatial information in a non-propositional format (Johnson-Laird 1983). Indeed, it appears that when readers read texts that include descriptions of spatial information, they construct representations of the information that preserve both stated and inferable spatial information in the form of spatial analogues rather than in the form of propositions (Haegeli et al. 1995; Glenberg et al. 1994; Morrow et al. 1987). One of the most intriguing questions to which neurocognitive methods should be able to contribute is how the brain handles non-linguistic spatial analogue information when it is explicitly or implicitly part of a text.

To illustrate the distinction between a situation model and a text base, Perrig and Kintsch (1985) provided two texts that describe a fictitious town. One was written to encourage a survey representation and the other was encouraged to provide a route representation, as illustrated in (7) and (8) respectively:

(7) North of the highway and east of the river is a gas station.
(8) On your left just after you cross the river you see a gas station.

Readers who read the survey text, including (7), were better at drawing a map of the town; readers who read the route text were better at remembering the text itself. The implication is that the survey text encouraged the readers to make a spatial situation model at the expense of a strong propositional text base; and, inversely, the route model allowed a strong propositional text base, but a less accurate situation model.

The focus on spatial analogues in text has served the goal of trying to demonstrate a distinction between texts and situations, a distinction not easy to verify so long as readers can use propositional information for both texts and situations. The logic of some of the research is to demonstrate that readers are sensitive to information in a way consistent with a spatial representation but not with a linguistic sequential representation. For example Morrow et al. (1987) had subjects read a narrative that brought the reader from one room to another in a building and answer probe questions about whether two objects were in the same or a different room. Readers' decision times were a function of the distance between the current room location (according to where the narrative was at the time of the probe) and the room that contained the pair of objects. A linear distance function is readily explained if the reader has constructed a model of the situation based on a walk through the rooms; it is less readily explained on the assumption that the reader built only an ordered list of text propositions.

Nevertheless, spatial models are only one possible realization of the idea of a situation model. In reading a narrative, one might expect time, rather than only space, to organize the reader's model of the situation. Indeed, Zwaan (1996) has demonstrated that readers use time phrases in a narrative—phrases such as an hour later or a moment later—to build temporal models. Reading times slowed down when there was a time shift (one hour) compared to when there was no time shift (a moment), and information after a time shift was more available than information before the time shift. More generally, Zwaan et al. (1995) argue that readers construct representations
of stories along five dimensions—time, space, protagonist, causality, and intentionality. Their event-indexing model assumes that events and the intentional actions of characters are the focal points of situation models, which are updated during reading along some or all of these dimensions.

6.6.2 Inferences build situations
The reader, in order to get from a text to a situation, generally must go beyond the explicit text information by making inferences. Because texts are never fully explicit, there are rich opportunities for the reader to add information and make inferences about what is in the text. Generally, the difference between a text base and a situation model is one of inferences. Text bases appear to be semantically 'shallow', that is containing meaning representations generated with minimal inferencing machinery, perhaps only those inferences needed to maintain referential coherence. Situation models, by contrast, are semantically deep, containing situation-specific meanings that generate rich inferences.

The issue in text research has been when do what kinds of inferences occur? Readers appear to generate those inferences that maintain referential coherence when the inference is textually required (Corbett and Dosher 1978; Dell et al. 1983; Haviland and Clark 1974; Just and Carpenter 1987). The theoretical approach to text meaning developed by Kintsch, including the Construction—Integration Model (Kintsch 1988), assumes that readers establish referential coherence across sentences by connecting pronouns to noun-phrase antecedents in building a new proposition. For example an encounter with a leads the reader to establish a proposition containing the immediately available antecedent of the pronoun or, if it is not available in the immediately available propositions, to search for an antecedent in the memory for the text base. Readers seek to make what they are reading referentially coherent, so either a pronoun or a noun without a clear referent triggers a process to establish co-reference with something available from the text representation. These co-referential inferences are considered minimal—needed not for an elaborated situation model but merely to make reading a text something different from reading a list of unrelated sentences. The question is what other kinds of inferences are made? And when?

Inferences that maintain causal coherence may also be made when needed (Keenan et al. 1984), although evidence is less strong on exactly when such inferences occur. On the other hand, for a whole range of elaborative inferences that a comprehender might be expected to make in establishing a situation model, the evidence is less strong. Most evidence does not support the assumption of early 'on-line' elaborative inferences (Corbett and Dosher 1978; McKoon and Ratcliff 1986, 1989, 1992; Singer 1979; Singer and Ferreira 1983), while some evidence suggests early inferences under restricted conditions (O'Brien et al. 1988). McKoon and Ratcliff (1989) suggest that inferences are encoded to different degrees of explicitness, with some, for example those that involve prediction of events, encoded only vaguely. Such less encoded inferences are readily made specific when required but are not specifically computed as part of comprehension. A related possibility is that elaborated inferences are not typically made as part of the text representation but can be observed in the situation model when readers are encouraged to attend to meaning (Fincher-Kiefer 1993). It is also possible that inferences required to maintain causal coherence among story elements are more likely to be made than other kinds of elaborative inferences (Trabasso and Suh 1993).

The possibility that all kinds of inferences are made routinely and automatically seems to have been clearly ruled out. Instead, it appears that readers are selective in drawing inferences, although the range of selectivity remains in question. The major theoretical issue in inferences can be captured by the contrast between the minimalist hypothesis (McKoon and Ratcliff 1992) and the constructionist hypothesis (Graesser et al. 1994). Whereas in the minimalist hypothesis only coherence-driven inferences are routinely made, with other inferences made only when task demands are right, the constructionist hypothesis is that a variety of inferences are made—those that address the reader's comprehension goals, those that explain why things occur, and those that establish global as well as local coherence.

It is difficult to draw clear conclusions about the inference question. Reader has to make some inferences just to keep the text base representation minimally coherent. And the reader is encouraged to make causal inferences just to keep the narrative structure (a situation model for a story) coherent. But other inferences, for example predictive inferences in which the reader might anticipate some event or some consequence of any action, are not required by considerations of coherence. Furthermore, they appear to require some effort and inferences of this 'forward' type, are especially subject to error. In fact, the search for robust inference drawing seems to be a search for active comprehension—understanding with learning—and the frustration of this search corresponds to the lament that many readers read passively, with minimal comprehension, rather than actively expending effort to connect and extend ideas in the text. To some extent, the large class of inferences that might separate the passive reader from the active reader can be viewed as under the control of individual reader and task variables. Thus, most readers can become more active readers, and hence inference generators, under the right circumstances. But under the mundane demands of typical experiments, there is little motivation to become an inference generator.

6.7 Individual differences in reading
The blueprint of the reader not only provides components for successful reading, it provides components that can fail and lead to problems in reading. When problems are relatively severe and occur without general intellectual problems, a child or an adult with reading difficulty is often categorized as having dyslexia, or specific reading disability. In many other cases, the reading problem may not lead to a label of specific reading disability. Although there may prove to be some important process deficits uniquely associated with specific reading disability, there are reasons at present to blur the distinction between specific and non-specific reading disability from a functional point of view. Individuals in these two categories seem to differ not so much in what their problem is, but in how they achieve in some other non-reading area. Indeed, the
evidence that the two groups' reading problems have different causes is weak (Stanovich 1988; Stanovich and Siegel 1994).

Figure 6.2, based on Fig. 6.1, represents the components of reading that one could imagine might go wrong. A reader could have trouble with processing printed inputs, or more generally visual inputs of certain kinds. A reader might know relatively few words, or not be able to use the needed meaning of a word, thus limiting lexical processes. A reader might have defective phonological processes, which would limit word identification and memory for words. On the comprehension end, a reader could have syntactic difficulties, have insufficient conceptual knowledge to support understanding of many texts, could fail to generate inferences, or could have a lower than normal working-memory capacity; any of these would limit the reader's ability to construct a good model of the text (text base) or of the situation, or both. There are additional processes in which the reader monitors comprehension, staying alert to coherence problems in the text itself and checking the mental model he or she is building against some criterion of sensibility.

These various possibilities are by no means equiprobable, and research has provided information on the source of comprehension problems. The next sections summarize this information.

6.7.1 Problems in lexical orthographic-phonological processes

Readers who fail to read words fail to comprehend. Thus word-level problems are potentially the most important in creating reading failures, because they would lead both to specific reading disability, which by definition entails problems with reading words, and to comprehension problems. The second of these problems results when obstacles to reading words create a bottleneck in comprehension (Perfetti and Lesgold 1979). More generally, because skill in comprehension rests in part on the efficient execution of lower-level processes within the constraints of working memory, verbal efficiency theory (Perfetti 1985) assumes that readers who lack efficient word-coding procedures are at risk for comprehension failure. Thus, even when words are accurately read, slow or effortful reading is often associated with comprehension problems. One specific interpretation of the verbal efficiency hypothesis is that the working memory bottleneck that produces reading problems is phonological (Shankweiler and Cain 1986).

When lexical processes are severely defective, the reading problem is manifest more directly, not merely in comprehension but in obvious word reading difficulty. The theoretical locus of a word reading problem is in the process by which orthographic and phonological information coheres to a word identity (see Fig. 6.2). Knowledge of letters and knowledge of phonemes that connect with letters in specific letter environments are necessary. A problem with either orthography or phonology produces lexical processing problems, and some research asks which of these sources is more important in any given case of reading failure. For example, in an ongoing study of familial aspects of dyslexia, monozygotic and dizygotic twins diagnosed as reading-disabled, are given tests that attempt to separate orthographic from phonological knowledge (Olson et al. 1990, Pennington 1990). For the phonological component, Olson et al. (1990) measure the time to read aloud pseudowords (both one- and two-syllable) for the orthographic component, they measure the time a reader takes to decide which of two homophonic letter strings is a real word, for example rain, rain. The assumption is that pseudoword reading reflects primarily phonological coding processes, whereas the word decision task requires knowledge about spelling patterns, independent of phonology. Both the phonological and orthographic component appear to have a significant heritable component (Olson et al. 1994).

The interdependence of orthographic and phonological knowledge is a difficult question. Research of Stanovich and West (1989) points to the possible independence of these two sources of knowledge, one primarily orthographic and one phonological.
7.2 Problems in Processing Lexical Meaning

and Understanding Inferences and Sentence Relations (from Understanding the Meaning of the Sentence, 1999)

In the 1960s, researchers were interested in how people process written language. It was believed that understanding the meaning of a sentence involves breaking it down into smaller parts: words, phrases, and clauses. However, recent research has shown that this process is more complex than previously thought.

According to the Activity Theory of Language, understanding a sentence involves not only breaking it down into smaller parts, but also incorporating the context in which the sentence is spoken or written. This means that the meaning of a sentence can change depending on the situation in which it is used.

For example, the sentence "He saw the man with the red umbrella" can have different meanings depending on the context. If the context is a rainy day, the sentence might mean that the man is carrying an umbrella to stay dry. If the context is a parade, the sentence might mean that the man is part of a performance.

This shows that understanding a sentence is not just a matter of breaking it down into smaller parts, but also involves incorporating the context in which it is used.

References:

related to another meaning of spade, it should have some initial activation. Indeed, both skilled and less skilled comprehenders showed longer times to reject ace (relative to control sentences) when it appeared 100 ms after the sentence. When ace appeared 850 ms after the sentence, however, only less skilled readers showed longer times to reject it. Other studies with this population (Gernsbacher and Faust 1990) suggest that the use of context is not a general problem for less skilled comprehenders. Rather, less skilled readers have a problem specifically with the suppression of irrelevant information.

Whether or not ineffective use of context is a source of reading problems has become a complex issue. Pre-scientific beliefs on this question seem to be that poor readers failed to use context in reading words. However, research on children's word identification led to the opposite result; less skilled readers use context in word identification at least as much and typically more than do skilled readers (see Perfetti 1985; Stanovich 1980, 1981). The understanding of this fact is that less skilled comprehenders are good users of context, which they rely on to identify words in compensation for weaker orthographic–phonological knowledge. However, Gernsbacher's book on meaning selection suggests a specific additional problem, one of using the semantic results of word identification.

6.7.3 Problems in processing syntax

Dyslexic children and even 'garden variety' poor readers—less skilled readers who do not necessarily meet all the criteria for specific reading disability—show problems handling various aspects of English morphology and syntax (Fletcher et al. 1981; Mann et al. 1984; Vogel 1973). The question is whether such problems, which are found across a wide age range, arise from some deficit in processing syntax or from some other source that affects performance on syntactic tasks.

Two hypotheses have been developed to account for problems observed in syntactic processing. On one account, syntactic problems, where they are observed, reflect a lag in the development of linguistic structures (Byrne 1981; Fletcher et al. 1981; Stein et al. 1984). Such an account, at first glance, seems consistent with the fact that syntactic deficits are observed in spoken language, not just written language. However, an alternative to a structural deficit was proposed by Perfetti and Lesgold (1977), who argued that the basic causes of less skilled comprehension were localized in working memory limitations. Problems in discourse comprehension, including syntactic problems, arise neither from intrinsic syntactic problems nor higher-level discourse structure problems, but from processing bottlenecks, partly, but not wholly, arising from lexical processing inefficiency.

There is now evidence that at least some syntactic problems do result from processing deficiencies. Crain and Shankweiler (1988; also Crain et al. 1990) have argued that, on the structural deficit hypothesis, one should expect differential problems in syntax; the problems encountered by less skilled readers should not be just the ones that also cause problems for skilled readers. For example, object-relative clauses, such as (9), are generally more difficult than subject-relative clauses such as (10), even for skilled readers.

9. The girl that the boy believed understood the problem.
10. The girl that believed the boy understood the problem.

Using data of Mann et al. (1984), Crain and Shankweiler (1988) showed that third grade skilled and less skilled readers produced the same pattern of errors rather than different patterns of errors on subject- and object-relative clauses. Crain et al. (1990) also report garden-path data that they argue are inconsistent with a structural deficit hypothesis, again relying on identical error patterns across groups of skilled and less skilled readers in support of this conclusion.

A more direct demonstration that the processing load affects syntactic performance was also provided by Crain et al. (1990). They found that problems with understanding temporal clauses (Push NP1 before/after you push NP2) increased as assumed processing load increased, either by making the NP more complex (not affecting the clause structure) or by not satisfying certain presuppositions associated with normal use of these clauses.

The possibility that reading problems and spoken language problems (aphasia) arise from processing limitations rather than structural deficits has gained increasing support from research with adults (Carpenter et al. 1994). This research has revealed a strong correlation between working memory capacity and language processing. Especially relevant for syntax are data reported by King and Just (1991) on subject and object relatives similar to those in (9) and (10). Not only did low-span readers (low in working memory for language) have more problems with object-relative sentences than did high-span readers (high in working memory for language), the problems, as indexed in reading times on words, were most severe where the processing load was hypothesized to be the greatest: at the second verb in the object-relative sentence (see (9)). The generalization here is that reading difficulties are localized in sentences at points of high processing demands (syntactic complexity, for example), and readers with lower processing capacity have problems at these locations. The problem is not syntactic deficits but the processing capacity to handle complexity.

Another perspective on this issue is helpful—the opportunity for practice. Because some syntactic structures are more typical of written language than spoken language, the opportunity for practice is limited by the ability to read. Thus, continuing development of reading skill as a result of initial success at reading and the parallel increasing failure as a result of initial failure—is undoubtedly a major contributor to individual differences in reading. Stanovich (1986) has discussed this rich-get-richer aspect of reading skill, borrowing, from Merton (1968) and Walberg and Tsai (1983), the Matthew metaphor: 'For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath' (Matthew XXV: 29). Seeing syntactic problems as Matthew effects places them intermediate to the syntactic deficit hypothesis and the processing capacity hypothesis. They can reflect either underdeveloped abilities—limited through restricted practice—or processing capacity—also limited through restricted practice with language.
6.7.4 Problems at the text level

What is the source of reading problems that are presented at the level of text comprehension specifically, as opposed to lexical or sentence levels? It is clear that at least some problems in comprehension can be explained by hypotheses that target lower level (lexical/phonological) processes within a limited processing-capacity working-memory system (Crain and Shankweiler 1988; Perfetti 1985; Carpenter et al. 1994). According to the blueprint of the reader, there are additional possibilities in the processing that builds mental models out of sentences. Readers must use knowledge outside of the text and apply inferences to the basic propositional content of the text, and these processes of knowledge use and inference generation can be the source of comprehension failure. (Note that simple lack of knowledge can also be the problem.)

The overall complexity of text comprehension (in terms of number of processes and possible interactions among them) implies corresponding complexity in the possibilities for comprehension failure. For example, some readers may be less likely to integrate sentences by establishing coreference, and, indeed, individual differences in coreferencing processes are found (Frederiks 1981). Or some readers may be less successful in monitoring their text comprehension and thus fail to build coherent and accurate mental models (Ryan 1982). Additionally, because some knowledge of text content is helpful in all reading, individual differences in knowledge can produce individual differences in comprehension (Anderson et al. 1977; Splich et al. 1979). However, low-knowledge readers can compensate for low knowledge with reading skill, at least to a limited extent (Adams et al. 1995). Readers who lack both content knowledge and reading skill have more serious problems; and the deleterious effect of low reading skill (and its motivational consequences) on learning through reading creates readers who lack knowledge of all sorts.

Among the several possibilities for problems, however, the one receiving the most attention is inferential processing. For example, Long and Golding (1993) conclude that skilled comprehenders are more able than less skilled comprehenders to make inferences about superordinate goals. More generally, Oakhill and Grahm (1988) summarize an array of evidence that suggests that less skilled readers fail to make a range of inferences in comprehension. Oakhill et al. (1997) conclude that this inference-making problem is general across spoken and written language, and even pictorial understanding. A general question of the hypothesis of specific comprehension deficits, including inferencing, is whether the observed differences in inferencing occur in the absence of problems in lower level abilities. Perfetti et al. (1996) suggest that although inferencing differences clearly have been established, they have generally not been established with procedures that assure the absence of lower-level lexical problems. Nevertheless, it appears that inferencing problems can arise for children who have no problems identifying words in context (Oakhill and Yuille 1986). Oakhill et al. (1997) further concluded that problems in inferencing observed in low skill readers are associated with reduced working memory capacity, providing an interesting integration of two lines of research that were initially quite separate. It seems safe to say that a specific inferencing disability, independent of other factors known to be critical for comprehension, remains unestablished.

In at least one study (Stothard and Hulme 1996), a careful attempt to delineate a group of less skilled comprehenders who show no problems in decoding and phonological processing has succeeded in showing distinctive types of comprehension problems: one type that occurs for children who also have decoding and phonological awareness problems and another type for those who do not have such problems. The latter group, which could be characterized as showing specific comprehension deficits, show problems in language processing generally and may be considered to have a comprehension-specific deficit, although not necessarily independent of processing limitations.

Another example of a specific comprehension problem is comprehension monitoring, found to be ineffective among less skilled readers in a number of studies (Baker 1984, 1985; Garner 1980). The test for comprehension monitoring is typically whether the reader can detect (explicitly refer to) an inconsistency introduced into a short text. Low-skilled comprehenders have been found to be especially poor at detecting higher-level text inconsistencies, those that would interfere with the construction of a coherent model of the text content (e.g. whether successive paragraphs are on unrelated topics). However, it is not completely clear whether these higher-level differences are independent of the reader's ability to construct a simple understanding of the text, for example at the proposition level; some evidence suggests that poor comprehenders fail to accurately represent or weight the propositions in a text (Otero and Kintsch 1992). The general interpretive problem here is that comprehension monitoring, like inference making, both contributes to and results from the reader's text representation. This makes it difficult to attribute comprehension problems uniquely to failures to monitor comprehension, as opposed to more basic comprehension failures.

6.8 Summary

Comprehending printed language shares many of the processing resources used in spoken language processes. A 'Blueprint of the Reader' entails a set of interrelated elementary processes and momentary representations. Printed word identification is the central, recurring event in reading that is not shared directly with spoken language processes. However, word reading builds on spoken language processes, specifically requiring phonological connections to be established to printed forms (letter strings and words). One of the most important conclusions arising from research in this field over the last 20 years is the importance of automatic phonological processes in the identification of printed words. Reading comprehension processes build on the identification of words, rapidly extracting context-sensitive meanings, assembling strings of morphemes into syntactic structures (parsing), building basic meaning units (propositions), integrating basic meaning units within and across sentences, and inferring additional information required to build a general (non-linguistic) representation of the content of a text. How these various components interconnect—where
interactions are unconstrained and where some information sources are privileged—is a major theoretical preoccupation of research on cognitive language processes, including reading. It is this architectural question, and in the individual differences that arise from different components of any candidate architecture, for which expectations are highest for increasing progress through cognitive and neurocognitive methods.

Acknowledgements

The author gratefully acknowledges Julie Van Dyke, Lesley Hart, Li Hai Tan, and Benjamin Xu for their input and assistance during the preparation of this chapter, which was supported in part by the Learning Research and Development Center of the University of Pittsburgh.

Notes

1. The term ‘phonological’ is widely used in the literature on reading, and will be used here. However, in most of its uses, what is referred to is the specification of the phonemes of a word, so ‘phonemic’ would be a more precise description. Note that ‘phonological’ can include suprasegmental information at the word level (syllabic stress in English; tone in Chinese) as well in phrasal units.

2. Whether semantic influences can operate quickly enough to affect the very beginnings of word identification is another matter. As reviewed in the section on word identification, the phonological activation of printed word forms appears to be so rapid and automatic, relative to semantic influences, that it seems better to say that meaning feedback operates as soon as it is available, but typically after some early cohering of phonological and orthographic information.

3. There are a number of proposals for how to conceptualize how parsing processes proceed and thus what causes garden-path phenomena. Minimal attachment remains the most general principle based on syntactic structures essentially independent of lexical factors. For other proposals within specific syntactic theoretical frameworks see Gibson (1991), Pritchett (1992), and Frazier and Clifton (1996). For a proposal outside the linguistic tradition with strong interactive assumptions see MacDonald et al. (1992). For a hybrid proposal within a syntactic framework, see Perfetti (1990). And for a review of the theoretical and empirical issues, see Mitchell (1994).

4. Assigning coreference, for example assigning pronouns to antecedents, is clearly the joint function of syntactic information and referential information from the discourse. Thus, as in the case of syntactic attachment, the communication between a sentence processing mechanism and a model of the discourse becomes an issue. Garrod and Sanford (1994) conclude that contextual information is immediately integrated during sentence processing, at least when the text provides strong grounds for the reader to commit a specific referential assignment to an ambiguous pronoun. Referential assignment and constituent attachment both illustrate necessary comprehension processes that require a discourse-sentence processing interface.
References


