56 Wang, R.F. *et al.* (1999) Mechanisms of reorientation and object localization by human children: a comparison with rats. *Behav. Neurosci.* 113, 475–485
57 Gouteux, S. and Spelke, E.S. (2001) Children's use of geometry and landmarks to reorient in an open space. *Cognition* 81, 119–148
58 Learmonth, A.E. *et al.* (2001) Toddlers' use of metric information and landmarks to reorient. *J. Exp. Child Psychol.* 80, 225–244
59 Hermer-Vazquez, L. *et al.* (2001) Language, space, and the development of cognitive flexibility in humans: the case of two spatial memory tasks. *Cognition* 79, 263–299
60 Hermer-Vazquez, L. *et al.* (1999) Sources of flexibility in human cognition: dual-task studies of space and language. *Cogn. Psychol.* 39, 3–36

Depth of processing in language comprehension: not noticing the evidence

Anthony J. Sanford and Patrick Sturt

The study of processes underlying the interpretation of language often produces evidence that they are complete and occur incrementally. However, computational linguistics has shown that interpretations are often effective even if they are underspecified. We present evidence that similar underspecified representations are used by humans during comprehension, drawing on a scattered and varied literature. We also show how linguistic properties of focus, subordination and focalization can control depth of processing, leading to underspecified representations. Modulation of degrees of specification might provide a way forward in the development of models of the processing underlying language understanding.

Anthony J. Sanford*
Patrick Sturt
Dept of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow, UK G12 8QB.
*e-mail: tony@psy.gla.ac.uk

Experimental research on human language processing has repeatedly demonstrated that the various sub-processes occur extremely quickly [1–3]. Perhaps because of the concentration on individual sub-processes, many researchers implicitly assume that these processes are not only initiated quickly, but are also completed quickly, resulting in the fast and dynamic construction of a fully articulated analysis of the linguistic input at each level. However, the time-course of processing and the depth of processing represent two orthogonal dimensions, and we currently know far more about the former than the latter.

In this article, we suggest that many processes are incomplete, and that interpretations are not as full as possible, but are often 'underspecified'. Recent work in computational linguistics and formal semantics has highlighted underspecification, which allows processing to proceed without maintaining a full analysis. Also, a variety of results in psycholinguistics show that language processing can be shallower than is commonly assumed. We shall argue that during comprehension, each word in a sentence does not necessarily contribute its full meaning, and these meanings are not always combined into higher-level phrase meanings through a fully determinate analysis.

Our case is that underspecification has a major role to play in the further development of process models of language comprehension. In fact, fully specified interpretations of language can often seem both undesirable and unnecessary. For instance, consider the pronouns in the following sentences:
(1) *Mary bought a brand new Hitachi radio.*
(2) *It was in Selfridge's window.*
(3) *Later, when Joan saw it, she too decided it would be a good purchase.*

A full specification of the referent of *it* in (2) is not possible. Did Mary buy the particular radio that was physically in the window, or was the one in the window just an exemplar of the set of radios? The interpretation of *it* in (3) offers even more possibilities. This example shows that processing might not occur to a fine grain. Indeed, it has been argued that there are many cases where the referent of a pronoun cannot be determined, and yet people are not concerned about this fact. Another example of an unresolved pronoun comes from the TRAINS corpus collected at the University of Rochester. The corpus is of dialogs about about train scheduling:
A: '*can we kindly hook up….uh…engine E2 to the boxcar at Elmira*'
B: '*okay*'
A: '*and send it to Corning as soon as possible please*'
B: '*okay*'

The pronoun *it* in the penultimate line is unresolved because it is ambiguous as to whether it refers to the engine, the boxcar, or both [4]. However, in this example, because the result of sending *it* to Corning remains the same, the referent of *it* may remain underspecified without affecting the key interpretation. The general question in these cases is whether or not a particular interpretation has been given, or whether some less specific representation has been formed.

Shallow processing in computational linguistics

Recent work in computational linguistics has used underspecified representations of text. So, although full parsing aims to recover a fully articulated grammatical structure for a sentence, 'shallow parsing' [5], aims simply to identify non-overlapping 'chunks' of structure in a text (Fig. 1). Choosing whether to build a full or shallow parser depends on how one wishes to use the resulting representations. For some tasks (e.g. automatic generation of indexes for large texts), shallow parsing is sufficient, whereas others (e.g. machine translation) require fuller analyses.

http://tics.trends.com 1364-6613/02/$ – see front matter PII: S1364-6613(02)01958-7

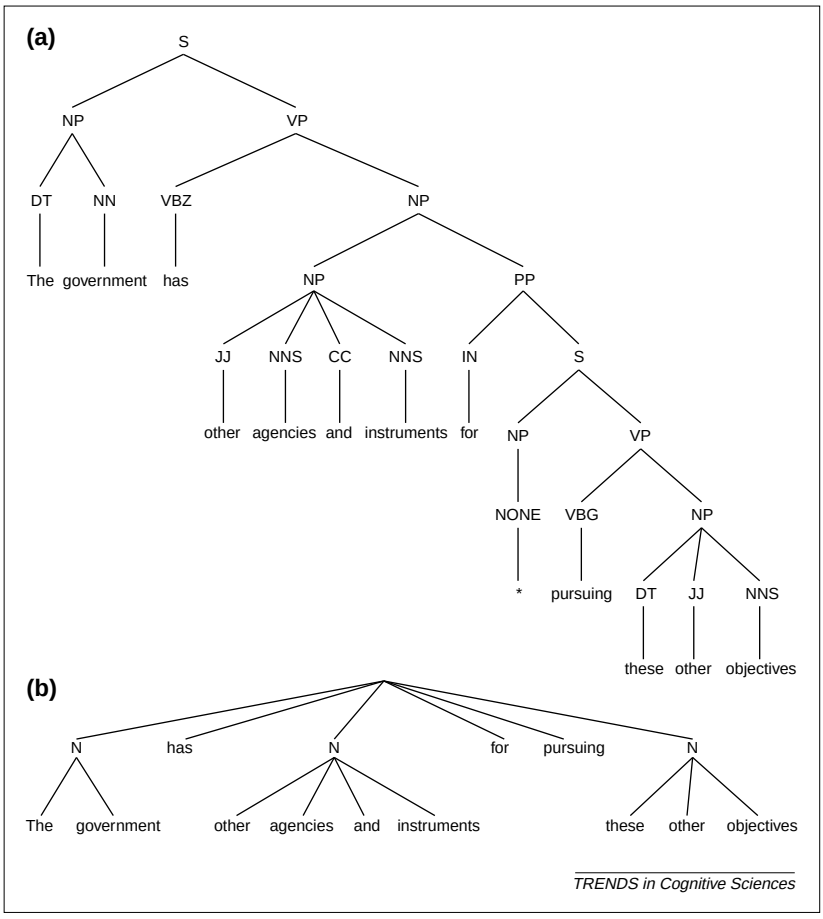

Fig. 1. Analyses of an example sentence from a large collection of syntactic trees for the Wall Street Journal [36]. Full parse (a), and shallow parse (b) of the sentence '*The government has other agencies and instruments for pursuing these other objectives*'. Full parsers construct an articulated grammatical analysis for the sentence. The parse tree (a) shows the hierarchical structure of the sentence constituents and their sub-constituents. By contrast, shallow parsers identify a subset of analyses for sentences. (b) shows the same sentence shallow-parsed into non-hierarchical noun groups, none of which have sub-constituents [5].

Other work has considered the use of underspecification to deal with ambiguity, especially with scope ambiguities, exemplified by '*Every kid climbed a tree*', meaning either there was some tree *x* such that every kid climbed *x*, or that for each kid *x* there was some tree which *x* climbed. Sentences with multiple quantifiers result in an explosion of ambiguity (Fig. 2).

Computational linguists have sought to bypass this explosion by representing sentence meanings that are neutral with respect to scope orderings, using underspecification (The 'Core Language Engine' [6] used 'quasi logical form' to achieve this). In machine translation between two related languages, it might not always be necessary to disambiguate a multiply-quantified expression fully [7].

Sometimes underspecified representations may be used for logical inference, without unpacking the scope-orderings of the ambiguity [8]. So, given '*At least one problem preoccupies every politician*', along with '*John is a politician*', one can infer that '*At least one problem preoccupies John*'. This inference can be made without committing to either scope ordering of the original sentence. An interesting question is whether such underspecified representations are used in human language processing.

Underspecification in human language understanding

There is growing evidence that underspecification does play a role in human processing. We will consider three cases: first, scope disambiguation, second, the selection of word meanings in cases of ambiguity of meaning, and finally evidence that word meaning is not always fully utilized in interpretation. We also describe evidence that reveals influences on the extent of interpretation, and claim that these could provide the foundation for a fuller theory of language comprehension than is currently available.

Consider quantifier scope first. If underspecified representations [8] are used by humans, then people should be able to integrate the meaning of a multiply-quantified sentence without committing to any one scope ordering. One recent study (Susan Tunstall, PhD dissertation, UMass, Amherst, 1998) compared the following:

(4) *Kelly showed every photo to a critic last month.*
(5) *The critic was from a major gallery.*
(6) *The critics were from a major gallery.*

There are two major readings for (4): in one, there could be just one critic, in which case continuation (5) should be read more easily. In the other, there is more than one critic, and continuation (6) should be read more easily. The results of a study of reading times showed that there was no preference, with (5) and (6) being read equally easily.

Work on word-meaning selection has also appealed to underspecification. To what extent do readers commit to a particular meaning of an ambiguous word? For instance, given '*John made his way to the bank*', do readers select between the financial institution and an earth-slope (river bank) before any text-driven reason for choosing [8,9]? In an eye-tracking experiment [10], radically ambiguous words like *pupil*, with two unrelated meanings (part of eye, or student), were contrasted with words like *newspaper*, having two related meanings; (printed paper, or company). For words like *newspaper*, people could delay their decision because the two meanings share features in common, which can serve as the word's contribution to the sentence-meaning. However, words like *pupil*, with unrelated features, could force an immediate commitment to a specific meaning, because there is no appropriate set of features compatible with both meanings that could be used in the interpretation. The eye-tracking data supported this claim. Similar claims have been made for the interpretation of metonymy [11]. Words like *Vietnam* can be interpreted literally (as a country) or metonymically (as a war). The evidence suggests that for well-established metonymies, a precise meaning is not established until needed. Given the frequency of such ambiguities, the results imply that people very

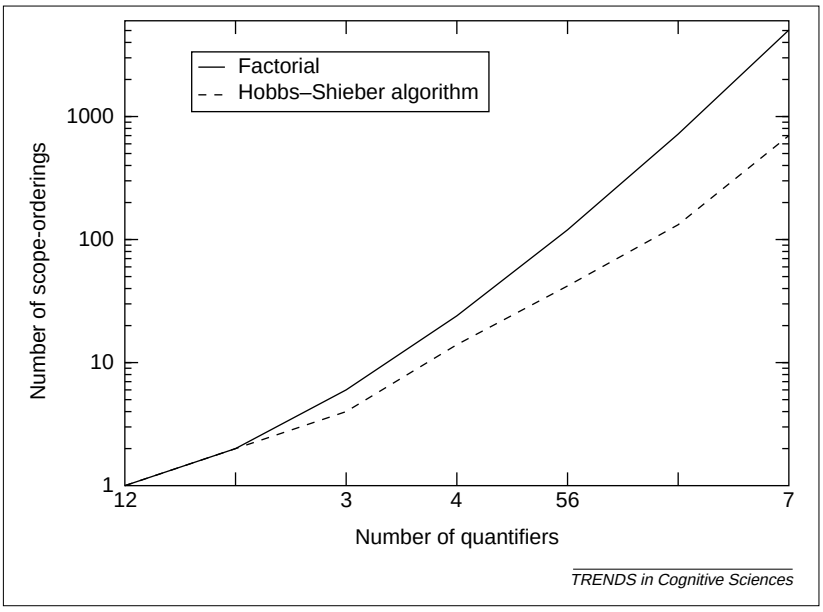

Fig. 2. Increase in ambiguity with number of quantifiers. Number of scope orderings for sentences with 1 to 7 quantifiers (1 quantifier: '*Every company folds*'. 7 quantifiers: '*Every department in most companies in many cities gives some businessman with few friends a pay-rise*'.). The graph shows both the number of possible mathematical permutations calculated naively (factorial function, solid line), and the number of logically permissible scope-orderings, calculated using the Hobbs–Shieber algorithm (dashed line) [37].

often use underspecified representations of word meaning. For reasons of experimental methodology, both of the experiments described above eventually disambiguated the ambiguous words one way or another. However, in real language, such disambiguating information might not always be provided, making it possible that hearers or readers never actually settle on a particular interpretation for some words.

Not noticing the evidence: the case of illusions

Another substantial, if patchy, literature provides dramatic evidence that the use of lexical semantics can be far from complete. Semantic anomalies are normally noticed. For instance, '*He drinks his tea with dog*' is obviously noticed, because the features of *dog* are inappropriate [12]. However, some anomalies are not noticed if the anomaly is similar in meaning to the 'right' word. In the so-called Moses illusion, many people answer 'two' to the question '*How many animals of each sort did Moses put on the Ark?*', instead of 'none' [13–16]. *Moses* is judged as being semantically similar to *Noah*, but less similar to *Adam*, and *Adam* is more easily detected if it is used to create the anomaly [14]. The Moses illusion is usually thought of as being part of the memory retrieval literature rather than the comprehension literature. However, similar failures to detect occur with plainly semantic mismatches that rely upon the recovery of word-meanings for their detection.

Barton and Sanford [17] tested the 'survivors' anomaly: to the question '*After an aircrash, where should the survivors be buried?*', half of the participants gave an answer like 'bury them where their relatives want'. They did not notice that *survivors* are living people. Thus, participants do not always incorporate all lexical meanings into their interpretations of sentences. In another test, subjects were asked questions about marriage laws in the UK, including '*Can a man marry his widow's sister?*' Only ~30% of participants noticed that to have a widow he would have to be dead!

People are clearly using shallow processing here. If the word *injured* is substituted for *survivors* in the 'survivors' problem, detection rates fall. The explanation is that the core meaning of *survivors* is 'to be alive after a life-threatening event', whereas for *injured* it is 'to have part of the body damaged' [17]. Being alive isn't part of the definition of *injured*, even though to say somebody was injured presupposes that they were alive at the time of referring to them. So, when semantic analysis based only on core meaning is incorporated into the interpretation, *survivors* is detected as anomalous, but not *injured*.

This argument explains why *survivors* is more readily detected, but not why detection is not 100%. The explanation depends on the fit of a word to the situation depicted in the text. Detection rates for *survivors* are much higher in a bicycle-crash than in the air-crash scenario, because the possibility of dead victims is much higher in an air-crash than in a bicycle crash [17]. So, when a word fits a situation, it can be processed less deeply than when it does not. Future experiments must decide whether it is recovery of word meaning, or integration into a final interpretation that is the locus of shallow processing.

Pragmatic and discourse influences on processing depth

The data above are consistent with general knowledge entering interpretation at an early stage [18]. Seldom-cited illustrations are the so-called Depth-Charge sentences [19,20], an example of which is '*No head injury is too trivial to be ignored*'. The default interpretation of this is 'No matter how trivial it might appear, a head injury should be treated', although grammar rules this out. Superficially it seems no different from '*No missile is too small to be banned*', which truly paraphrases into 'No matter how small it might be, a missile should be banned'. However, it is a straightforward step to see that the head injury case paraphrases to 'No matter how trivial a head injury appears to be, it should be ignored'.

These effects can be attributed to the role that general knowledge plays in interpretation. The default is that trivial head injuries still require investigation, and that is how the Depth-Charge sentence is understood. But to reach this interpretation, local semantic analysis must be incomplete or overridden by more global constraints. The same can be said of the '*widow's sister*' example: because it makes sense to ask whether a man can marry his wife's sister if his wife dies, this pragmatic interpretation is produced

Box 1. Text change detection

The problem with underspecified representations is that of determining what grain of representation is being used. In reading, surface wording is quickly forgotten, the gist being retained [a]. Re-reading a text in which a word has been changed to one that is at the same level of granularity as the original should result in a failure to detect the change. However, a substantial change in meaning should be detectable. Consider example (A) (taken from Ref. [b]):

(A) *The newsagent had just hired a new paperboy to cover the downtown area. The paperboy* ***finished*** *his rounds after he ate his breakfast. There were a lot of deliveries to be made.*

Hence, on the above argument, the change *finished* → *completed* should be less detectable than the change *finished* → *started*. Here, the critical word *finished* falls in the main clause of the sentence. In (B) it is in the subordinate clause position:

(B) *The newsagent had just hired a new paperboy to cover the downtown area. After the paperboy* ***finished*** *his rounds, he ate his breakfast. There were a lot of deliveries to be made.*

Detections should be poorer in (B) than in (A).

The results of an experiment using texts of this type conformed to this prediction [b]. Texts changed every 6 s, and first-change detections were more likely the larger the change in semantic distance, and if the change was in the main clause (Fig. I). (See also [c] for a related eye-tracking experiment.)

References

a Sachs, J.S. (1967) Recognition memory for syntactic and semantic aspects of connected discourse. *Percept. Psychophys.* 2, 437–442

b Sanford, A.J. (2002) Context, attention, and depth of processing during interpretation. *Mind. Lang.* 17, 188–206

c Raney, G.E. and Rayner, K. (1995) Word frequency effects and eye movements during two readings of a text. *Can. J. Exp. Psychol.* 49, 151–171

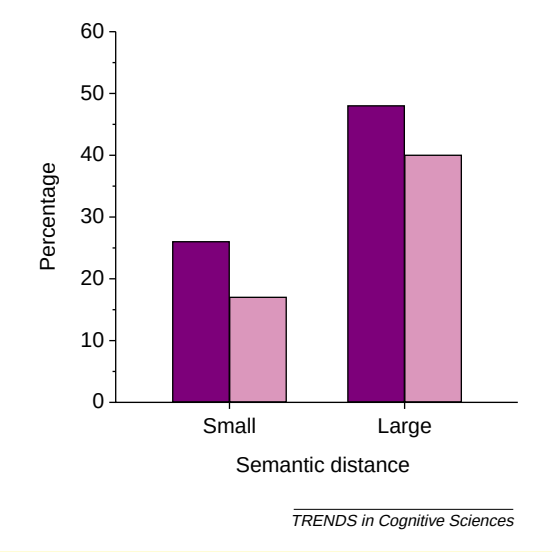

Fig. I. Percentage correct detections of text changes in the main clause of a sentence (purple bars), and subordinate clause (pink bars). as a function of semantic distance. Redrawn with permission from Ref. [b].

(at the expense of local semantics). The term 'Pragmatic Normalization' has been used to cover cases where a knowledge-based interpretation is given to sentences expressing unusual propositions, such as '*If you don't break the rules, then I'll tell the authorities*' [21,22–25].

Focus and subordination also play a role in extent of processing. Cleft sentences manipulate focus, indicating what question a sentence answers. So, '*It was John who answered the door*' answers the question '*Who answered the door*?' The Moses anomaly is easily detected in cleft sentence constructions: '*It was Moses who put two of each kind of animal on the ark. True or False?*' [26]. Subordination also relates to focus. It has been shown that false assertions in logically subordinate clauses (7) are less likely to be noticed than they are in main clauses (8) [27]:

(7) *The liver, which is an organ that is found only in humans, is easily damaged by alcohol.*

(8) *The liver, which is an organ that is easily damaged by alcohol, is found only in humans.*

There have been few experimental studies of these issues. However, recently, we have been investigating the use of a text change-blindness procedure, an analogue of the method in visual perception in which observers see a scene, which, after exposure, changes in some way [28]. Failures to detect the changes are diagnostic of a failure to attend to the appropriate aspect of the scene. An example of the reading analogue is given in Box 1.

Conclusions

Underspecification has been a growing concern in formal and computational linguistics. We have presented evidence from several disparate literatures showing that underspecification occurs during human language understanding [29], (see also [30]). Until now, there has been relatively little exchange between disciplines on these issues although underspecification could easily be considered within most approaches (there are hints in both connectionist approaches [31], and symbolic accounts [32]). We believe that it will be shown to play an important, and probably central role in future process models.

Underspecified representations are more tolerant in that they are not so likely to be disturbed by changes in interpretation forced by subsequent linguistic input. Given the fact that natural speech deviates greatly from most notions of strict grammaticality, such tolerances are essential. Indeed, although we have concentrated on reading in this article, it is quite possible that underspecification is greater with spoken language, given the fact that the rate of information uptake is not under the control of the listener, resulting in time pressures.

Underspecification also fits the fact that discourse has a thread: it is about some things, and not about

other things [33]. The task of the producer is to cause the receiver to process the thread, and not irrelevant ramifications. Subordination and focus represent ways of signalling what is important in a discourse, and we have seen how they play a role in determining depth of processing. It is an empirical question, for instance, whether reference resolution can be determined at a finer grain in main clauses, or for focussed entities, than for subordinate, or unfocussed ones. Signalling importance is a general issue, and a theory of discourse comprehension must address this and its relation to processing. Another means of signalling importance is through focalization in narrative [34]. In narratives with more than one character, it is typical to find that the perspective of one character (the main character, or a narrator) is adopted in the text. There is evidence that inferences concerning the main character are made more easily than ones concerning a secondary character [35]. This is because the narrative thread is carried by the main character, and we would expect processing to be more shallow with respect to secondary characters.

Exploring how underspecification is used in relation to the task demands of language processing is a challenge to current theories of comprehension, and the development of new techniques, such as text-change blindness, should provide the tools to take up the challenge.

Acknowledgement
We thank ESRC for support through grant R000223622.

References

1 Altmann, G.T.M. (1999) Thematic role assignment in context. *J. Mem. Lang.* 41, 124–145
2 Altmann, G.T.M. and Kamide, Y. (1999) Incremental interpretation at verbs restricting the domain of subsequent reference. *Cognition* 73, 247–264
3 Tannenhaus, M.K. *et al.* (1995) Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634
4 Poesio, M. and Reyle, U. (2001) Underspecification in anaphoric reference. *Proceedings of 4th International Workshop on Computational Semantics*, Tilburg University
5 Ramshaw, L.A. and Marcus, M.P. (1995) Text-chunking using transformation-based learning. In *Proceedings of 3rd Workshop on Very Large Corpora* (Yarovsky, D and Church, K., eds), Somerset, New Jersey, Association of Computing and Linguistics
6 Alshawi, H., ed. (1992) *The Core Language Engine*, MIT Press
7 Alshawi, H. *et al.* (1992) Swedish–English QLF translation. In *The Core Language Engine* (Alshawi, H., ed.), pp. 277–309, MIT Press
8 Reyle, U. (1993) Dealing with ambiguities by underspecification: construction, representation and deduction. *J. Semantics* 10, 123–179
9 Copestake, A. and Briscoe, T. (1995) Semi-productive polysemy and sense extension. *J. Semantics* 12, 15–68
10 Frazier, L. and Rayner, K. (1990) Taking on semantic commitments: processing multiple meanings vs multiple senses. *J. Mem. Lang.* 29, 181–200
11 Frisson, S.P. and Pickering, M.J. (1999) The processing of metonymy: evidence from eye movements. *J.Exp. Psychol. Learn. Mem. Cogn.* 25, 1366–1383
12 Kutas, M. and Hillyard, S.A. (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205
13 Erickson, T.A. and Matteson, M.E. (1981) From words to meaning: a semantic illusion. *J. Verbal Learn.Verbal Behav.* 20, 540–552
14 Van Oostendorp, H. and De Mul, S. (1990) Moses beats Adam: a semantic relatedness effect on a semantic illusion. *Acta Psychol.* 74, 35–46
15 Van Oostendorp, H. and Kok, I. (1990) Failing to notice errors in sentences. *Lang. Cogn. Process.* 5, 105–113
16 Hannon, B. and Daneman, M. (2001) Susceptibility to semantic illusions: an individual differences perspective. *Mem. Cogn.* 29, 449–462
17 Barton, S.B. and Sanford, A.J. (1993) A case-study of anomaly detection: shallow semantic processing and cohesion establishment. *Mem. Cogn.* 21, 477–487
18 Sanford, A.J. and Garrod, S.C. (1998) The role of scenario-mapping in text comprehension. *Discourse Process* 26, 159–190
19 Wason, P. and Reich, S.S. (1979) A verbal illusion. *Q. J. Exp. Psychol.* 31, 591–597
20 Natsopoulos, D. (1985) A verbal illusion in two languages. *J. Psycholinguist. Res.* 14, 385–397
21 Fillenbaum, S. (1971) Processing and recall of compatible and incompatible question and answer pairs. *Lang. Speech* 14, 256–265
22 Fillenbaum, S. (1974) Pragmatic normalization: further results for come conjunctive and disjunctive sentences. *J. Exp. Psychol.* 102, 574–578
23 Schlesinger, I.M. (1968). *Sentence Structure and the Reading Process*, Mouton.
24 Christianson, K. *et al.* (2001) Thematic roles assigned along the garden path linger. *Cogn. Psychol.* 42, 368–407
25 Ferreira, F. *et al.* (2001) Misinterpretations of garden-path sentences: Implications for models of reanalysis. *J. Psycholinguist. Res.* 30, 3–20
26 Bredart, S. and Modolo, K. (1988) Moses strikes again: focalization effects on a semantic illusion. *Acta Psychol. (Amst.)* 67, 135–144
27 Baker, L. and Wagner, J.L. (1987) Evaluating information for truthfulness: the effects of logical subordination. *Mem. Cogn.* 15, 247–255
28 Simons, D.J. and Levin, D.T. (1997) Change blindness. *Trends Cogn. Sci.* 1, 261–267
29 Sanford, A.J. (2002) Context, attention, and depth of processing during interpretation. *Mind. Lang.* 17, 188–206
30 Ferreira, A. *et al.* (2002) Good enough representations in language. *Curr. Dir. Psychol. Sci.* 11, 11–15
31 McClelland, J.L. *et al.* (1989) Sentence comprehension: a parallel distributed processing approach. *Lang. Cogn. Process.* 4, 287–335
32 Frazier, L. and Clifton, C. (1996) *Construal*, MIT Press
33 Grimes, J.E. (1975) *The Thread of Discourse*, Mouton
34 Toolan, M.J. (1988) *Narrative: A Critical Linguistic Introduction*, Routledge
35 Sanford, A.J. *et al.* (1998) The influence of character type on processing background information in narrative discourse. *Mem. Cogn.* 26, 1323–1325
36 Marcus, M.P. *et al.* (1993) Building a large annotated corpus of English: the Penn Treebank. *Comput. Linguist.* 19, 313–330
37 Hobbs, J.R. and Shieber, S.M. (1987) An algorithm for generating quantifier scopings. *Comput. Linguist.* 13, 47–63