Cognitive Science 28 (2004) 1–44

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Interpretation-based processing: a unified theory of semantic sentence comprehension

Raluca Budiu*, John R. Anderson[1]

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received 20 February 2002; received in revised form 6 August 2003; accepted 2 October 2003

Abstract

We present interpretation-based processing—a theory of sentence processing that builds a syntactic and a semantic representation for a sentence and assigns an interpretation to the sentence as soon as possible. That interpretation can further participate in comprehension and in lexical processing and is vital for relating the sentence to the prior discourse. Our theory offers a unified account of the processing of literal sentences, metaphoric sentences, and sentences containing semantic illusions. It also explains how text can prime lexical access. We show that word literality is a matter of degree and that the speed and quality of comprehension depend both on how similar words are to their antecedents in the preceding text and how salient the sentence is with respect to the preceding text. Interpretation-based processing also reconciles superficially contradictory findings about the difference in processing times for metaphors and literals. The theory has been implemented in ACT-R [Anderson and Lebiere, The Atomic Components of Thought, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 1998].

Keywords: Language comprehension; Semantics; Computational modeling; Metaphor; Moses illusion; Text priming; ACT-R

Ambiguity is one feature of human language that often frustrates the attempts to automatize its understanding by computers: not only can words have multiple meanings, but sometimes the meaning of a word is not taken at face value. Everyday language is often nonliteral; figurative devices such as irony, indirect request, metaphor, metonymy, or hyperbole are common and are understood easily. Metaphor is a particularly pervasive device: it is a rich source of new words (recent examples include *web* and *couch potato*) and, moreover, according to researchers such

* Corresponding author. Tel.: +1-412-422-7214.

E-mail addresses: ralucav@cs.cmu.edu (R. Budiu), ja@cmu.edu (J.R. Anderson).

0364-0213/$ – see front matter © 2003 Cognitive Science Society, Inc. All rights reserved.
doi:10.1016/j.cogsci.2003.10.001

as Lakoff (1987), Lakoff and Johnson (1990) or Reddy (1993), language is often shaped by existing, conceptual metaphors.

Humans comprehend language not only in the presence of ambiguity or nonliterality, but also in the presence of noise. Real-time communication is inherently noisy—often, the communication medium is imperfect (e.g., a bad connection on the phone or a discussion in a noisy room) and people make errors in pronunciation or choice of words, but their communication partners are able to grasp the gist of their message. Sometimes speakers say what they did not intend (for example, when they commit slips of the tongue), but we still understand them. Not only do listeners often recover from mispronunciations or slips of the tongue, but sometimes they are unable to notice them in a sentence. For instance, when asked *When an aircraft crashes, where should the survivors be buried?*, about 80% of people do not detect the anomaly (i.e., that the survivors need not be buried—Barton & Sanford, 1993). Even if they are warned in advance that the sentence may be distorted, about 40% of participants still do not notice the inconsistency in a statement such as *Moses took two animals of each kind on the ark*, in spite of knowing that *Noah*, rather than Moses, is the character of the ark story in the Bible (Erickson & Mattson, 1981). (This phenomenon is called *Moses illusion*.) These facts suggest that ignoring minor discrepancies in communication is such a basic feature of our language system that we cannot easily turn it off.

In this article we argue that metaphor comprehension and lapses in detecting semantic inconsistencies are facets of the same mechanism of language processing; that we understand metaphors easily for the same reasons for which we fail to notice semantic distortions. We propose a new theory of sentence understanding, called interpretation-based processing (INP) that hinges on the concept of prior knowledge. INP postulates that the same processes are involved in the processing of both literal and nonliteral language and shows that literality is only a matter of degree. Although a word may seem inappropriate, a rich sentence context that contains a lot of known information can often help people identify what the sentence is about and make them grasp the intended meaning of that word. Moreover, if the sentence context precedes the "wrong word" (be it metaphor, semantic distortion or even a poorly chosen literal), then people can get what the word refers to without even detecting that it was inappropriate or not used literally. That is, the literal meaning of a word can be bypassed if the other preceding words in the sentence are informative enough. This theory naturally explains priming effects due to context: rich contexts can provide interpretations for sentences; these interpretations, in turn, can facilitate the processing of new meanings.

One consequence of our theory is that it reconciles the contradictory findings in psycholinguistic studies that have compared the comprehension speed of metaphors and literals: whereas some of these studies did not find any significant difference in overall processing time for metaphoric and literal sentences (Budiu & Anderson, 2002; Inhoff, Lima, & Carroll, 1984; Keysar, 1989; Ortony, Schallert, Reynolds, & Antos, 1978; Shinjo & Myers, 1987), others found an advantage of literal sentences over metaphoric ones (Budiu & Anderson, 2001; Gibbs, 1990; Onishi & Murphy, 1993).

INP is embodied as an ACT-R (Anderson & Lebiere, 1998) model that actually processes sentences. (Further on we use INP to refer to both the theory and the model.) ACT-R is a cognitive architecture that has served as a framework for successfully modeling a large variety of problem-solving and memory tasks (see http://www.act-r.psy.cmu.edu/ for a list of articles that describe ACT-R models). Implementing the theory in ACT-R has a number of advantages, perhaps the most important of them being that it allows the model to perform in real time tasks described in the psycholinguistic literature. Thus, the responses produced and the time taken by an ACT-R model for a task can be compared directly to the human participants' data from the same task. We use this testing methodology to evaluate our model. The commitment of ACT-R to real-time processing also creates another challenge for INP: like humans, it must perform all of its processing at the speed of only a few hundred milliseconds per word. This constraint proves to be a severe test for any theory.

INP simulates several datasets—three of them come from the metaphor literature (Budiu & Anderson, 2002; Gerrig & Healy, 1983; Onishi & Murphy, 1993), two datasets involve the Moses illusion (Ayers, Reder, & Anderson, 1996; Reder & Kusbit, 1991), and one is concerned with text priming (Schwanenflugel & White, 1991). Although there are psychological theories that separately address metaphor processing, semantic illusions, text priming, real-time sentence processing, or that have been realized as running models, ours is the first to simultaneously achieve all of these constraints. Moreover, to the best of our knowledge, INP is the first full-fledged domain theory for metaphor understanding that reconciles those contradictory empirical findings regarding the processing times of metaphors and literals. INP also relates metaphor processing to other text-processing empirical phenomena that were not traditionally regarded as connected to metaphors and shows that failures of the language-processing mechanism (e.g., Moses illusions) reflect the same process that enable language flexibility in the comprehension of metaphors. As such, we believe that INP constitutes another step of progress in coming to an understanding of human language processing.

In the rest of the paper we discuss the general behavior of our model, then we describe the experimental tasks to which it has been applied. Finally, we comment on other empirical predictions that INP makes and, in the conclusions section, we discuss how our model compares to other theories of sentence processing.

1. The interpretation-based—processing theory

INP is implemented in ACT-R (Anderson & Lebiere, 1998), a general theory of human cognition. In Appendix A we review those general ACT-R concepts and mechanisms that play an important role in our INP model; here we present the core concepts of INP—interpretation and background knowledge and, also, its main representational and process assumptions.

1.1. *Interpretation and background knowledge*

The task of the INP model is to produce syntactic and semantic representations for the input sentence and to relate the sentence to prior (or background) knowledge, which may contain facts such as *College students live in dorms*, *Noah took two animals of each kind on the ark*, or *The night sky is filled with stars*, or other more specific propositions derived from text preceding the input sentence. Theoretically, the relationship with the prior knowledge could be quite complex: one could imagine a variety of inferences being drawn about the new sentence (e.g., what caused that sentence, what its consequences may be, how typical it is, whether it contradicts other knowledge). But, practically, this multitude of inferences is not feasible in a system such as ACT-R, in which each step of processing takes at least 50 ms.[2] Therefore, INP is limited to drawing only one such "inference"—namely, finding a known fact that overlaps most with the current sentence. We call this fact *the interpretation*. It is the position of the interpretation in the overall map of prior knowledge that enables the model to perform more complex inferences, when (and if) they will be needed. Budiu (2001) and Budiu and Anderson (2000) showed that this minimalist inference mechanism is powerful enough to capture many aspects of sentence memory. Specifically, Anderson et al. (2001) illustrated how the sentence interpretation (which they call referent) can help or interfere with later memory of the text. In this article, we show that tentative interpretations based on sentence fragments can help or interfere with lexical processes or even with the comprehension process itself.

INP incrementally builds the representations and tries to find an interpretation as it "reads" the words in the sentence, before reaching the end of the sentence. The aim of the model is to "guess" the interpretation of the current sentence as soon as possible. The incrementality of language comprehension (i.e., that people do not wait for the end of the sentence to process the incoming words) is supported by a number of experimental studies (Marslen-Wilson, 1973, 1975; Oakhill, Garnham, & Vonk, 1989; Traxler, Bybee, & Pickering, 1997; Tyler & Marslen-Wilson, 1982) another indirect demonstration is found in one of the metaphor studies discussed in this article (Gerrig & Healy, 1983).

When a sentence communicates old information, its interpretation naturally corresponds to a prior-knowledge proposition that matches it (almost) perfectly, that is to a proposition that contains the same (or highly similar) concepts in the same roles. For a sentence such as *At the restaurant the man paid the waiter*, such an interpretation can be the fact *The customer paid the waiter*, which is part of our prototypical restaurant knowledge. Alternatively, when the sentence is novel, INP needs to be content with an interpretation that matches only part of the input sentence. For instance, *Tengiz Abuladze directed "Repentance"* is an obscure sentence for many people; a possible interpretation could be *The person directed a play*. ("Repentance" is actually a movie.) Whereas in the case of sentences transmitting known information the interpretation can be regarded as the "meaning" of the sentence, for novel sentences, the interpretation is rather a "hook" or an "anchor" point in the prior knowledge: it does not necessarily match all the concepts in the sentence, although it shares some.

1.2. *Representation*

Recent ACT-R models (e.g., Anderson, Bothell, Lebiere, & Matessa, 1998; Anderson, Budiu, & Reder, 2001; Salvucci & Anderson, 2001) have developed a fragmented style of representation, in which the information about a single event or object is spread among several chunks (i.e., pieces of declarative knowledge). INP applies this style for both syntactic and semantic (or propositional) representations, which are closely related to those used by Anderson et al. (2001). Here we focus on the semantic representation[3]. Fig. 1 depicts the semantic representation corresponding to the sentence *The college students were taught by professors of good reputation*.

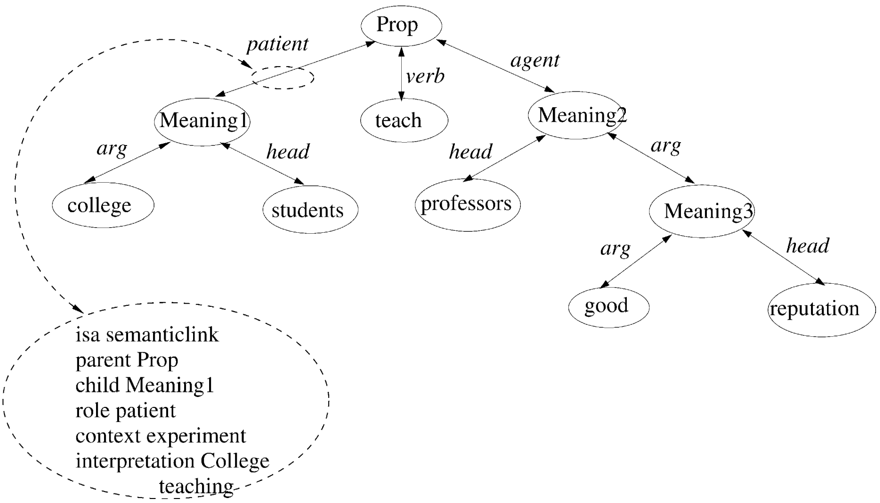

Fig. 1. Semantic representation for the sentence *The college students were taught by professors of good reputation.*

The semantic structure of the sentence is encoded as a tree; the nodes in this tree correspond to meanings. The root of the tree represents the meaning of the whole sentence; the interior nodes are composite meanings (e.g., the meaning of the noun phrase *college students*); and the leaves are meanings of individual words. The key assumption is that each of the nodes and edges in this tree is represented as a declarative chunk in ACT-R. This assumption enables the model to reason about each piece of the meaning separately in its incremental processing of the sentence. Fig. 1 shows one example chunk that encodes the fact that *Meaning1* is the patient of the proposition *Prop*. Note that the edges contain information about the nodes they connect, about the role of the child within its parent, about the context in which they occur, and also about the interpretation of the current sentence.

1.2.1. Semantic similarities

The ultimate purpose of our model is to account for phenomena at the semantic level of language. To achieve this purpose, it relies on the concept of *semantic similarity*, which drives the process of activation spreading. According to ACT-R, when an item comes in the focus of attention, it spreads activation to other items to which it is associated. The amount of activation spread depends on the strength of association, which, in turn, in the case of INP, depends on the semantic similarity.[4] (This dependence is linear and described in Appendix B.) In our model, words just read get in the focus of attention and spread activation to other semantically similar facts in the prior knowledge.

The question that remains to be answered is how INP computes semantic similarities. Since we do not have a theory of semantic similarity, we use an existent theory—Latent Semantic Analysis (LSA—Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998)—to obtain similarities between different words; then, we define similarities involving more complex structures (e.g., propositions) based on LSA word similarities. Thus, the model needs to receive as input similarities between words. Appendix B specifies the rules for calculating similarities of complex structures and how these similarities determine the strengths of associations in the ACT-R model and drive the spreading of activation.

As mentioned, we use LSA to define the basic similarities between words. LSA is a mathematical technique that generates a semantic space starting from a text corpus. It represents word meanings as vectors in a high-dimension space; dimensions in this space are different texts and the meaning vector reflects how frequently the word occurs in each dimension text. Then LSA applies singular value decomposition to reduce the dimensionality of the semantic space. The similarity between two meanings is calculated as the cosine of the angle between the corresponding vectors in the smaller-dimension semantic space. This technique can be extended to compute the similarity between a word and a passage. LSA was used to simulate many psycholinguistic phenomena such as vocabulary acquisition (Landauer & Dumais, 1997), emergence of natural categories (Laham, 1997), predication (Kintsch, 2001), and metaphor comprehension (Kintsch, 2000). Although LSA may not always offer a perfect definition of similarity (even for some of our simulations), it is clearly quite successful and provides a solid, independently defined constraint for our model. It could be replaced by more reliable definitions of similarity (e.g., participant ratings) if these were available. The qualitative predictions of the model depend on similarity orderings rather than on exact values obtained from LSA or from another theory.

1.3. Process

INP has two components: a syntactic component and a semantic one; they are both available on line as a single model at the *Published Models* link at http://www.act-r.psy.cmu.edu/. The syntactic component processes each new word in the sentence as it is “read” and builds both the syntactic and the semantic representation for the input sentence. The focus of this article is on the semantic component. For the sake of completeness, we describe the syntactic processes that INP carries out in Appendix C, although we do not make much reference to them in the rest of this article. The existence of a syntactic component is the main difference between INP and the purely semantic model in Budiu (2001) and shows that comprehension of basic sentences can be achieved rapidly in a production system, even when both syntactic and semantic processes are accounted for. Given that we want to show that INP can run in real time, it is critical that the model also deal with syntax.

The semantic component of INP attempts to guess what the sentence is about as it reads it. It retrieves a proposition (a *candidate interpretation*) from the prior knowledge that best matches what the model has read and, then, it checks whether the candidate interpretation is validated by subsequent information; if it is not, the model attempts to find another candidate. The semantic processor acts on each semantic unit (or on what it believes to be a semantic unit); roughly speaking, semantic units correspond to verbs, adverbs, and noun phrases. Thus, the search or validation of the interpretation does not occur on auxiliary verbs or prepositions or even on every component noun of a noun phrase, although the syntactic processor acts on each of these when it builds the semantic representation of the input.[5] For instance, instead of looking for an interpretation after each word in the noun phrase *the college students*, the semantic processor waits until after INP read all three words; if the noun phrase were *the college students with child dependents*, the semantic processor would act after reading *students* and, then, finding that the phrase continues, it would act again on *dependents*. Thus, the model makes a trade-off between the tendency to wait until it is sure that the noun phrase ended and a need to keep only a limited number of items unprocessed by the semantic processor. Specifically, it keeps at most one head noun in store before triggering the semantic processing for that noun phrase.

Let us go through how INP processes a sentence such as *The college students were taught by lecturers* (Fig. 2). We emphasize here only the actions of the semantic processor; Appendix C contains a more detailed discussion of how the model builds the syntactic and semantic representations shown in Fig. 2. After reading the words *the college students*, which together form

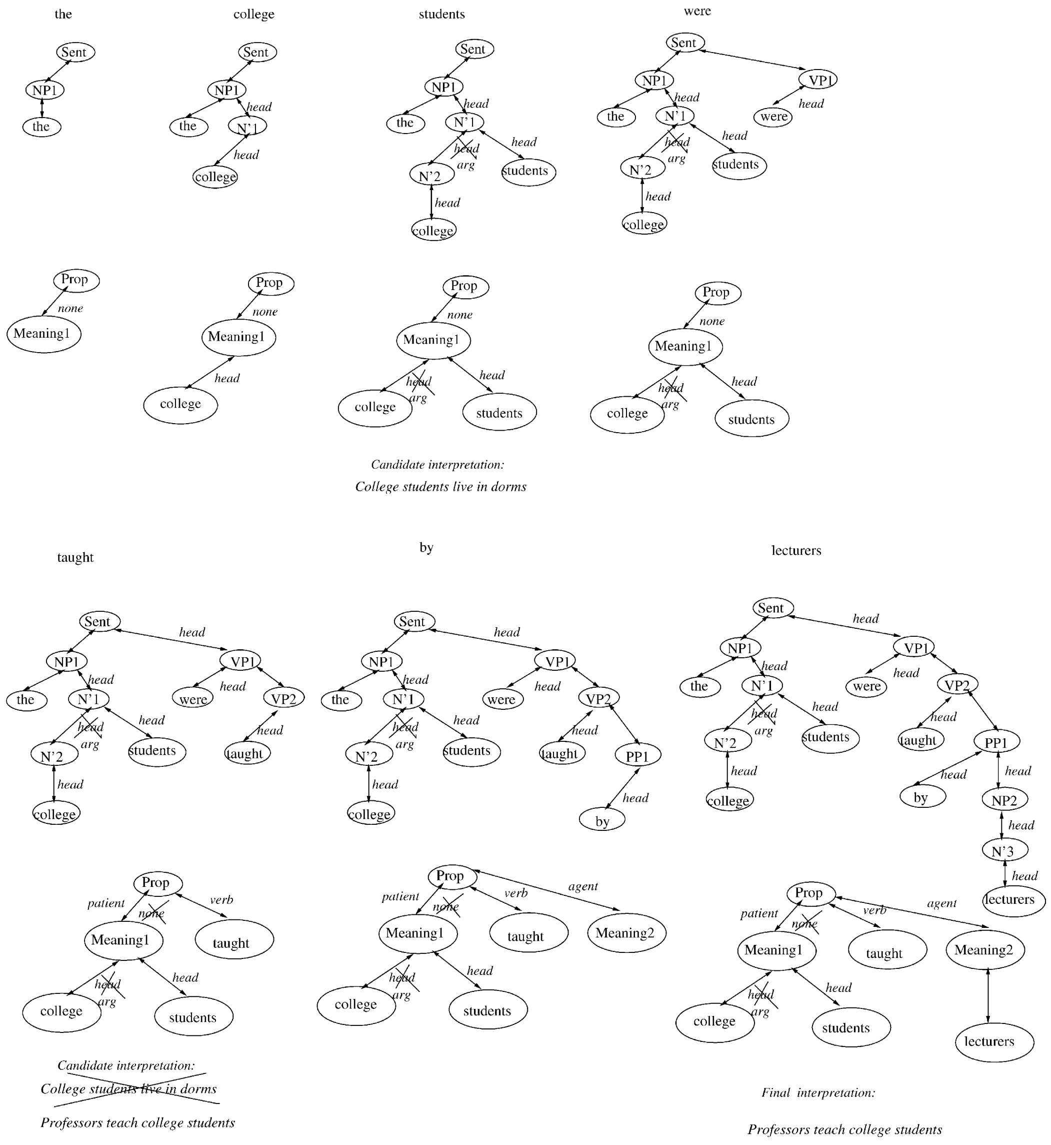

Fig. 2. Processing of the sentence *The college students were taught by lecturers*.

a noun phrase with the meaning represented by the chunk *Meaning1*, INP attempts to guess an interpretation for the input sentence based just on the words read. Thus, it looks for a fact in the background knowledge that involves the composite meaning *Meaning1* (i.e., *college students*) and makes that fact its current candidate interpretation (e.g., *College students live in dorms*). Next, when the main verb *taught* is read, the model checks whether the current word phrase matches the verb in the current candidate interpretation; in our case it does not, so INP needs to search for another interpretation involving *Meaning1* and *taught* and in which *taught* is a verb. Let us assume that INP selects *Professors teach college students*; that fact becomes our current candidate interpretation. This candidate interpretation is validated by the last word, *lecturers*, provided that the meanings of *lecturers* and *professors* are similar enough.[6]

Fig. 2 illustrates the two ways in which our model deals with transient syntactic ambiguity. One is by postponement. Thus, the model does not assign the patient role to *college students* until enough of the sentence has been processed to indicate that this assignment is appropriate. The other method used for transient ambiguities is to revise incorrect commitments that may have been made on the basis of existing evidence. So, for instance, *college* is initially assigned to a head role and then this role is changed. Our model does not have a full treatment of all possible transient ambiguities any more than it has a treatment of all possible syntactic structures. However, in both cases we think that the existing treatment can be extended and that it shows how syntactic ambiguity can be addressed in our framework, rather than a full-fledged, completely sound theory of syntactic ambiguity. The syntactic processing in the model is discussed in more detail in Appendix C.

The flowchart in Fig. 3 summarizes the behavior of the semantic processor. When it is invoked for the first time, INP has no semantic interpretation and attempts to find one that involves the current word phrase (e.g., *the college students*). Once it found a candidate interpretation, it matches that interpretation against subsequent word phrases. If at some point INP fails to find

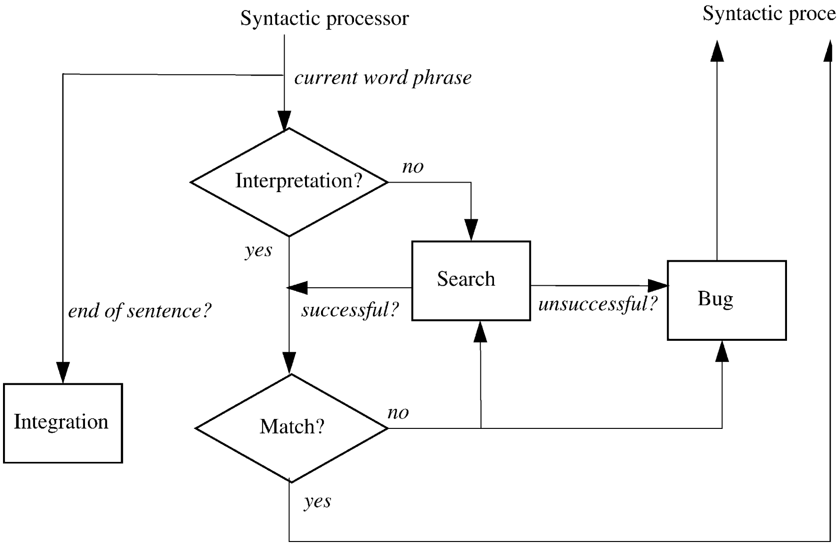

Fig. 3. The behavior of the semantic processor.

an interpretation for the current sentence, it records this failure event by creating a special chunk called *bug*. At the end of the sentence, the model integrates the semantic representation of the read sentence with that of the found interpretation. Next we discuss the steps in Fig. 3 in more detail.

1.3.1. Search

The search process (Box Search in Fig. 3) tries to find a proposition in the prior knowledge that best matches what was read from the input sentence. INP selects the prior-knowledge proposition that has the highest activation above the retrieval threshold. If there is no such proposition, the semantic processor returns control to the syntactic processor. Before doing so, it records the failure by creating a *bug* (Box Bug). Otherwise, if the search is successful, INP checks whether the current word phrase matches the found proposition (Box Match); if they do not match, the model either goes back to the syntactic processor or continues the search by looking at the next best proposition (Box Search). The decision to stop the search is probabilistic. If INP decides to move to the next word without having found an interpretation, it creates a *bug* chunk (Box Bug), which registers the failure to find an interpretation and some extra information about the context in which the failure occurred (e.g., current word phrase, current role, previous candidate interpretation). Bugs are bumps in the comprehension process and INP uses them to keep track of the local failures; later in the sentence, when more information is available, the model may try to recover from those failures. A smooth, bug-free comprehension indicates that the sentence is very close to some already known fact. Thus, bugs are useful in verification tasks, to decide whether the sentence is true or false.

To match the speed of human comprehension, INP must perform a very efficient search—it cannot spend time selecting many bad candidates and then rejecting them. Rather, it should find the right candidate interpretation as soon as possible in the search process: each failure costs time (each time INP finds a wrong interpretation it must spend extra time to look for another candidate interpretation). The order in which interpretations are selected is determined by their activation. The information that INP has gathered about the sentence spreads activation to the interpretation. At each moment, INP keeps the meanings of the last three word phrases processed in the focus; these meanings should occur in the correct interpretation of the sentence and, therefore, they should be highly similar and, hence, strongly associated to that interpretation.

1.3.2. Matching

We saw that a candidate interpretation is accepted only if it matches the meaning of the current phrase (Box Match). Matching compares the current word phrase with the concept in the same role in the interpretation—for instance, if the current word phrase is a verb, it compares it to the verb in the interpretation. In INP, matching is based on similarity and is realized through activation spreading. If the current word is similar enough to the corresponding concept in the interpretation, the activation spread from the current word, which is in focus, will increase the overall activation of the concept above a threshold[7] and the interpretation will match. Otherwise, if the word and the concept in the interpretation are not similar enough, the activation of the concept will remain under the threshold.

We should emphasize the distinction between search (Box Search) and match (Box Match): why match if we know that the immediately preceding search was successful? The search does not take into account the thematic roles of the word phrases; it returns any proposition involving the last three word phrases. The match step makes sure that the current word phrase actually matches the proposition concept in the same role. This strategy corresponds to the result of Ratcliff and McKoon (1989) that relational information is accessed only later in the comprehension process.

1.3.3. Interpretation priming

During the search process (Box Search), beside the last three meanings processed, INP keeps in focus the candidate interpretation. The candidate interpretation remains in focus for a short while even after it has been invalidated and, while in focus, influences the search for an interpretation by spreading activation to other facts similar to itself. Because it matched some meanings that may be no longer in focus, the previous candidate interpretation represents a synthetic memory of the sentence. However, INP discards it relatively soon to avoid getting trapped into trying only propositions that are similar to it.

1.3.4. Integration

The basic assumption behind the integration phase (Box Integration) is that the cognitive system spends some time at the end of the sentence (or clause) to coherently relate the current sentential input to prior (episodic or permanent) knowledge. Thus, integration in INP is conceptually similar to Just and Carpenter's (1980) wrap up at the end of the sentence.

With each new word that it processes, INP builds a part of the semantic representation (i.e., a new semantic link) for the input sentence; each of these parts contains information about the current interpretation. This information reflects INP's current "belief" of what the sentence is about. As INP goes through the entire sentence, that "belief" (i.e., the current interpretation) may change, so that, at the sentence, the final interpretation may differ from previous candidate interpretations that were encapsulated in the various parts of the representation. Therefore, INP needs to revise those parts that are no longer correct. This process is termed *integration*.

Let us go once more through the example from the beginning at this section and discuss it in more detail. Fig. 4 shows how the semantic processor works on the initial part of the sentence *The college students were taught by lecturers.*

First, it processes the meaning *Meaning1* of the noun phrase *the college students* and, with that concept in focus, starts looking for an interpretation. All propositions containing similar concepts get an activation bonus (arrows in Fig. 4). The proposition with the highest activation is picked up (assume it is *College students live in dorms*) and promoted as the current candidate interpretation. Next, the semantic processor processes the meaning of the word *taught*, which is known to be a verb. INP verifies whether the meaning of *taught* matches the verb of the current candidate interpretation; because *live* and *taught* are dissimilar, too little activation spreads from *taught* (in focus) to the concept *live* in the dorm proposition. Therefore, that proposition will be invalidated and another candidate proposition sought. This time, both *Meaning1* and *taught* (together with the dorm proposition) are in focus, so they spread activation towards similar propositions. The activation spread from the sources combines additively, such that the proposition that is most similar to all the chunks in the focus gets retrieved. Let us assume

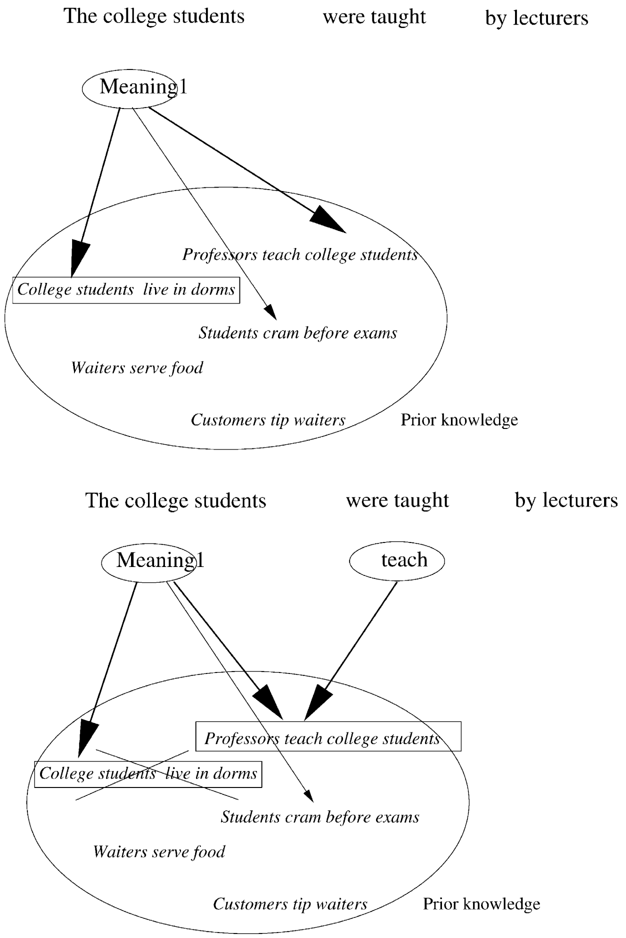

Fig. 4. Semantic processing of the initial part of the sentence *The college students were taught by lecturers*. Arrows represent spreading activation; the width of the arrow is proportional to the amount of activation.

that this proposition is *Professors teach college students*. INP checks whether its verb matches the current word *teach*; because it does, INP accepts this proposition as a current candidate interpretation. Next, the meaning of *lecturers* is processed; the model knows that this word is an agent (being introduced by the preposition *by*), so it matches it against *professors*, the agent of the current candidate interpretation. Assuming that the meanings of *lecturers* and *professors* are similar enough, there will be enough activation spreading from the input meaning to *professors* so as to make it raise above the threshold. Therefore, the candidate interpretation will be validated and accepted as a final interpretation. At the end of the sentence, during the integration phase, the model makes sure that all the components of the semantic representation contain correct information about the final interpretation *Professors teach college students*.

To summarize, our semantic model is a very simple search-and-match process. Although all steps in Fig. 3 take time, the search is the most expensive part—the number of repetitions of this step influences significantly the time spent for comprehending the sentence. Candidate interpretations proven invalid may help with finding new candidates. At the end of the sentence, the integration phase makes sure that the structures created are consistent. Bug chunks keep track of local failures of comprehension and they can be used to make various decisions. The model is non-deterministic—decisions are probabilistic and the activation-based mechanisms through which the steps in Fig. 3 are implemented are noisy (see Eq. (A.2) in Appendix A, for a description of activation noise in ACT-R).

2. Experimental validation

In this paper we show that INP can reproduce a number of critical results in the literature. We attempt to reproduce both the qualitative trends in the data and, at least approximately, the quantitative values. To reproduce the quantitative effects we estimated several parameters. Given that these experiments tend to report relatively few data values and that the model depends on relatively many parameters, one may think that INP could be coerced to predict any results. Therefore, we strive to show that the model predicts the qualitative effects and could not predict opposite effects. Although our principal interest is in the qualitative effects, given the constraints in ACT-R on rate and success of processing, it is not obvious that we would be able to reproduce the actual times and accuracies. Therefore, we thought it important to show that INP could also produce the quantitative results.

There are three types of parameters that occur in our simulations: (a) general ACT-R parameters, (b) word similarity parameters, and (c) parameters influencing the control structure of INP (i.e., probabilities for various branches of the flowchart in Fig. 3).

There are only two ACT-R parameters that we estimate in our simulations. One of these is the threshold parameter, τ (see discussion for Eq. (A.2) in Appendix A), which determines how active a chunk has to be in order to be retrieved. The other is a latency factor, F (see Eq. (A.3) in Appendix A) that scales activation into retrieval times. For these parameters the ACT-R theory does not stipulate any values; they vary widely among different ACT-R models (e.g., Anderson et al., 1998). Table 1 shows the values of these parameters for each simulation.[8]

To set the similarity parameters for one simulation, we averaged the LSA distances (Landauer & Dumais, 1997) over all materials in the corresponding experiment. For instance, the similarity between a literal and a metaphoric word was instantiated to the average LSA distance between the literals and metaphors used in all the trials of the experiment. One exception to this rule was the text-priming simulation—for this simulation, we did not have access to all the materials in

Table 1
Latency factor (s) and retrieval threshold for various simulations

Experiment	Retrieval threshold	Latency factor
Gerrig and Healy (1983)	−1.750	0.060
Onishi and Murphy (1993)	−3.500	0.003
Budiu and Anderson (2002)	−3.250	0.003
Ayers et al. (1996), Reder and Kusbit (1991)	−0.350	0.060
Schwanenflugel and White (1991)	−0.550	0.080

the original experiment (Schwanenflugel & White, 1991), so we based our LSA estimate on a single example. The use of a single average similarity value in the simulations is possible because the quantitative predictions of the model are a monotonic function of similarity (i.e., the higher the similarity, the higher the effects predicted by the model). We discuss the actual values for the similarities, as well as the remaining class of parameters, when we present each experiment.

Beside these three kinds of parameters, all our simulations use the value of 150 ms for the time to read a word; it reflects perceptual processes involved in reading, which we do not model here. The timings of our simulations are largely determined by this parameter and by each production cycle taking 50 ms. Although the setting of the latency factor can stretch or shrink these timings a little, our models are committed a priori to the rough times for processing sentences in all these tasks and it is not a trivial accomplishment that these models get in the ballpark of human performance.

2.1. *Metaphor comprehension*

Perhaps the best domain for illustrating INP is metaphor comprehension. Language is rich in metaphors. Almost all text contains metaphors of which neither the reader nor the author are aware (Gibbs, 1992). The ease with which we often process metaphorical language illustrates how comprehension requires the sort of semantic stretching that is at the core of INP. Whereas people show considerable facility with common and practical metaphors, it is less clear how they process novel metaphors. Many studies have compared the processing of sentences with novel metaphors with that of literal sentences and the results have been contradictory. Thus, Ortony et al. (1978), Inhoff et al. (1984), Shinjo and Myers (1987), Keysar (1989), Budiu and Anderson (2002) showed that, when the metaphor is preceded by a rich and supportive context, the metaphoric sentence can be understood as fast as a literal one, whereas Gibbs (1990), Onishi and Murphy (1993), Budiu and Anderson (2001), reported experiments in which participants were slower to read or verify metaphoric sentences compared with literal sentences. Giora (1997) proposed the graded-salience theory of metaphor comprehension, which states that the ease of comprehension is controlled by how salient the referent of the metaphor is. Salience depends on factors such as supportiveness of preceding sentence context and goodness and/or familiarity of the metaphor. Other models (see Gibbs, 2001 for a review) have also proposed that context and familiarity of metaphor both play a role in metaphor comprehension.

Next, we discuss simulations for three metaphor comprehension experiments: Gerrig and Healy's (1983), Onishi and Murphy (1993), and Budiu and Anderson (2002). Gerrig and Healy (1983) experiment shows that the sentence context preceding the metaphor can facilitate its comprehension; the other two experiments compare the processing of various metaphoric sentences with similar literal sentences.

2.1.1. *Metaphor position: Gerrig and Healy (1983)*

Gerrig and Healy (1983) showed that the position of the metaphor within a sentence may influence the speed of comprehension. They presented their participants with two kinds of sentences: sentences starting with a metaphor (e.g., *Drops of molten silver filled the night sky* in which *drops of molten silver* are a metaphor for *stars*, *The parallel ribbons were followed*

Table 2
Mean reading times (s) for metaphorical sentences from Gerrig and Healy (1983): data and model

Type of sentence	Reading times	
	Data	Model
Metaphor-first	4.21	4.21
Metaphor-last	3.53	3.70

by the train in which *parallel ribbons* are a metaphor for *tracks*) and sentences ending with a metaphor (e.g., *The night sky was filled with drops of molten silver*, *The train followed the parallel ribbons*); one type of sentence was usually obtained by making the other passive.

Gerrig and Healy measured reading times for these kinds of sentences and found that participants read metaphor-first sentences more slowly than metaphor-last sentences. Another experiment in the same study established that this result was not an artifact of the different sentence structure of the two types of targets. A related result has been reported by Peleg, Giora, and Fein (2001): they show that at the beginning of the sentence metaphoric words may be processed as literals initially, whereas at the end their metaphoric meaning may be accessed immediately.

Table 2 presents the reading times of the metaphoric targets in the two conditions from the first experiment in Gerrig and Healy (1983). This result is a nice demonstration that people dynamically interpret and reinterpret the sentence as they read it. If they waited until the end to assign an interpretation to the sentence, there should be no difference between the two conditions. The existence of a difference supports a key assumption of INP: incrementality.

Let us take a look at how the semantic component of INP behaves on Gerrig and Healy’s sentences. First, we consider metaphor-first sentences, such as *Drops of molten silver filled the night sky*. The first words (*Drops of molten silver filled*) suggest that the sentence may be about a container holding liquid silver, but the final words (*night sky*) do not match such an interpretation. Therefore, the model must reject the container interpretation and find a new candidate interpretation, which could be the correct interpretation *Stars fill the night sky*, provided that *stars* and *drops of molten silver* are similar enough. But switching from the container interpretation to the stars interpretation costs INP extra time. On the other hand, such a switch happens less often in the case of metaphor-last sentences. For a metaphor-last sentence such as *The night sky was filled with drops of molten silver*, it is more probable that, after reading *The night sky was filled with*, the model select the correct stars interpretation. The stars interpretation would be next validated by the last words of the sentence (*drops of molten silver*). Thus, INP predicts that metaphor-first sentences take longer than metaphor-last sentences, because the former require rejecting one interpretation and replacing it with another one. Table 2 (third column) shows the latency results produced by INP.

A critical assumption is that the topic and the vehicle of the metaphor (e.g., *drops of molten silver* and *stars*) are semantically similar. The similarity should be high enough to ensure that, once the entire sentence is read, INP will find the right interpretation, given the support of the rest of the sentence. The value of this similarity can influence the model’s latency and

Table 3
Variation of INP's predictions as a function of the similarity between the metaphoric word and its referent for the Gerrig and Healy (1983) simulation

	Similarity					
	0.18	0.28	0.38	0.48	0.78	1.00
Latency (s)						
Metaphor-first	4.26	4.22	4.18	4.11	4.05	3.74
Metaphor-last	3.83	3.70	3.66	3.64	3.63	3.61
Error rate						
Metaphor-first	0.45	0.10	0.01	0.00	0.00	0.00
Metaphor-last	0.46	0.01	0.00	0.00	0.00	0.00

Note. Latency stands for average latencies on targets for which the model is able to find the correct interpretation.

ability to find the right interpretation. To obtain the results reported in Table 2, we set the similarity between the metaphoric concept and its corresponding literal (e.g., *drops of molten silver* and *stars*) to 0.31. This value is the average of all LSA distances (Landauer & Dumais, 1997) between the metaphoric phrases and the literal phrases in the targets used by Gerrig and Healy (1983). Table 3 shows how the predictions of INP vary for other similarity settings. High values of similarity wash away the difference between metaphor-first and metaphor-last sentences: if the metaphoric word is very similar to the literal, the chance of getting the final interpretation before reaching the last semantic unit (e.g., *the night sky*) is considerably increased for metaphor-first sentences. The extreme similarity value of 1 corresponds to the case of a literal sentence. On the other hand, small similarity values make it harder for INP to find the right interpretation, as shown by the higher error rates and by the increased latencies for both kinds of targets. These results map nicely onto Genter and Wolf's (1997) empirical finding that the similarity between the two terms of the metaphor affects the speed of comprehension of the metaphor.

Critical to the predictions of the model is the significantly smaller chance of an interpretation switch for metaphor-last sentences. The model's basis for capturing the latency pattern in Gerrig and Healy (1983) is that metaphor-first sentences are reinterpreted once more at the last concept (*the night sky*), whereas metaphor-last sentences do not need a reinterpretation in most cases. This difference is a consequence of the character of knowledge in the long-term memory: if there are few propositions matching *The night sky* or *The night sky was filled*, then there will be a high chance that the right interpretation for the sentence *The night sky was filled with drops of molten silver* is found before the last concept (*drops of molten silver*) and no reinterpretation will be necessary. On the other hand, if there are many propositions matching the beginning of that sentence, it is possible that a reinterpretation occur. However, one can show that, under reasonable assumptions, the contents of the knowledge base does not affect the basic result that metaphor-first sentences are understood more slowly than metaphor-last sentences. Appendix D reproduces a proof from Budiu (2001) that these predictions hold in general and do not depend on the particular values of the parameters or on the contents of the background knowledge.

2.2. *Metaphors versus literals*

From Table 3 we can draw one important prediction of INP with respect to the processing of metaphors versus literals: sentences that end with a metaphor should be almost as easy as equivalent literal sentences (compare the case when the similarity is 1, corresponding to the literal, with the other cases), whereas sentences that start with a metaphor should be harder than sentences that start with an equivalent literal. The next two experiments both study these predictions and compare literal and metaphor processing.

These experiments involve reading of a text and, then, processing of a target sentence. The target sentence contains a metaphoric or literal reference to a concept from the text. To simulate these experiments, INP relies on the same kind of processes for both metaphoric and literal targets; the only difference is that metaphors are less similar to their referents than literals are.

2.2.1. *Onishi and Murphy (1993): anaphoric versus predicative metaphors*

In the Gerrig and Healy's (1983) simulation we saw that INP predicts different reading time patterns for predicative and anaphoric metaphors, when compared with literals. The model predicts, as found by Onishi and Murphy (1993), Peleg et al. (2001), Shinjo and Myers (1987) that predicative metaphors (e.g., D.G. Rossetti's metaphor *A sonnet is a moment's monument*) should not have much of an effect when compared with literals. Typically, the predicate is new information in either case and, in itself, may not help too much in the process of relating the sentence to the background knowledge; however, in general, the preceding sentence context, if informative enough (as in the case of metaphor-last sentences in the simulation of Gerrig & Healy, 1983), manages to compensate for this problem and connect the current sentence to the prior knowledge. This connection to background knowledge, once established, can, in most cases, smoothly accommodate the new information contained in the predicate. However, as found by Gibbs (1990) and Onishi and Murphy (1993), there should be a deficit in the processing of anaphoric metaphors because full interpretation of the sentence is "delayed" by the metaphor. Here we focus on modeling the study by Onishi and Murphy because they explicitly compared predicative and anaphoric metaphors.

Onishi and Murphy (1993) showed their participants stories like those in Table 4. The stories could contain either a literal or a metaphoric target. In one experiment, the target was anaphoric (left column in Table 4) and in another experiment, the target was predicative (right column in Table 4). Like Gibbs (1990), they found a difference between reading times for metaphoric and literal anaphoric targets; however, they found no difference for the case of predicative targets. The reading times in the two experiments are presented in Table 5.

This simulation is the first in this article that involves comprehension of sentences containing new information (henceforth called *novel* sentences). When simulating this study, INP tries to relate the novel sentence to information in the preceding text. The model attempts to find an interpretation for the read sentence among propositions from the passage. If, at some point, it finds an interpretation (which may be later rejected), that interpretation must share with the input at least some information. This interpretation can, thus, constitute a hook for relating the novel sentence to the context during the integration phase. Note that any partially matching interpretation (be subsequently rejected or the final one) can be such a hook. Ideally, the hook should be the proposition that matches best the input sentence (i.e., the final interpretation, if

Table 4
Sample materials from Onishi and Murphy (1993)

Anaphoric		*Predicative*
Felicia was a feline fanatic, who had two persians and a siamese. The siamese was her favorite, and she treated her like a child. One day it would not eat its food, though Felicia tried to coax it. After babying it for an hour, to no avail, she became worried. She called Joseph, her usual veterinarian, for advice. He was well aware of Felicia's doting attitude towards her pets. Felicia described her problem with her siamese.		Felicia was a feline fanatic, who had two persians and a siamese. The siamese was her favorite, and she treated her like a child. One day it would not eat its food, though Felicia tried to coax it. After babying it for an hour, to no avail, she became worried. She called Joseph, her usual veterinarian, for advice. He was well aware of Felicia's doting attitude towards her pets.
	Targets	
"My princess won't eat", she informed him.	[*metaphoric*]	"The cat is my princess", she informed him.
"My cat won't eat", she informed him.	[*literal*]	"The cat is my favorite", she informed him.
	Ending	
Joseph said, "Bring her in, there's an open slot at noon."		Joseph said, "Bring her in, there's an open slot at noon."

any, or, otherwise, the candidate last rejected). In Haviland and Clark's (1974) terms, the hook is the "given" part of the input. If no hook is found during semantic processing, the integration cannot be performed. The more "given" information in a sentence, the higher the chance that it will be related to the preceding text. Sentences with metaphors offer fewer opportunities for finding a hook than equivalent literal sentences, because the metaphors are less similar than the literals to their antecedents in the text.

INP captures the basic results using the same processes for both metaphors and literals. In the case of anaphoric targets, the difference in reading times between metaphoric and literal targets is due to the different similarities between the critical word (metaphor or literal) and its antecedent. At the beginning of the sentence, when the anaphoric metaphor is read, most often the model fails to find an antecedent for it and pays a time cost for that failure; however, as it reads more words, those help to relate the current sentence to the passage and the model ends up with an interpretation of the whole sentence. For anaphoric literal sentences, the model is more successful in finding an antecedent for the literal, so the comprehension is smoother and faster. There is no deficit in the processing of predicative metaphors relative to predicative literals because both involve new information and because the previous sentence

Table 5
Mean reading times (ms) for metaphoric and literal targets in Onishi and Murphy (1993): data and model

	Data		Model	
	Anaphoric	Predicative	Anaphoric	Predicative
Metaphors	2262	2146	2271	2301
Literals	1912	2054	1877	2330

context facilitates the finding of an interpretation for the sentence, in which the new information can be accommodated.

For this experiment, we needed an estimate of the similarity between the metaphor or literal and their referent in the text. We used the average LSAs between metaphors/literals and their intended antecedent in the story, as described by Onishi and Murphy in Appendix A of their study. The values obtained were 0.12 for metaphors, 0.3 for literals in the anaphoric experiment. In the predicative experiment, there were two types of literals—those that played the subject role in both metaphoric and literal targets (e.g., *cat* in the sentence *The cat is my princess/favorite* from Table 4) and those that played the predicate role in the literal targets only (e.g., *favorite* in the same sentence). The average LSAs between these literals and their antecedents in the story were 0.65 for the subject literals and, respectively, 0.41 for the predicate literals.

The LSA values seemed very small, especially for the first experiment (possibly because of difficulties in assessing the correct antecedent—see the discussion in the section on Budiu & Anderson, 2002)—for instance the literal *stomach* was rated at 0.10 LSA distance from the antecedent *pregnant woman's belly*, whereas the metaphor *barrel* was at 0.17 distance from the same antecedent. Therefore, we decided to add a constant increment of 0.15 to all LSA values. This correction also brought the LSA values for the anaphoric part of this simulation closer to values used in the other simulations.

2.2.2. More on anaphoric metaphors: Budiu and Anderson (2002)

The study by Ortony et al. (1978) is often cited as showing that even anaphoric metaphors do not result in a comprehension deficit, when compared with literals. Budiu and Anderson (2002) report a word-by-word reading study that examined their findings in more detail. In our experiment, participants read passages like those in Table 6, followed by a target sentence that could either be literal or metaphoric[9] and then judged the truth of a probe sentence. Ortony et al. (1978) reported no difference in the comprehension times for metaphoric versus literal target sentences, but did not collect word-by-word reading times.

Table 6
Sample materials from Budiu and Anderson (2002)

During history seminars, a massive young man always yawned and never paid any attention to the discussions. He was a very good linebacker who had been all-state in football. The seminar always came after his training sessions, so he was very tired.		Every year the Localville Women's Society for Animal Protection has a meeting. They bring in snacks, eat, and report about what was accomplished during the year. But this year, a major discussion topic was the new city regulations that allowed people to buy live animals from ethnic food stores.
	Targets	
The bear hibernated in class	[*metaphoric*]	The hens clucked noisily
The athlete slept in class	[*literal*]	The women talked noisily
	Probes	
The man dozed during the class	[*true*]	The ladies discussed loudly
The man daydreamed in class	[*false*]	The ladies sang loudly

Table 7
Reading times (ms) from Budiu and Anderson (2002) and model results

	Data		Model	
	Met noun	Lit noun	Met noun	Lit noun
Noun + verb RT	1237	1168	1229	1186
Ending RT	778	794	841	886
Sentence RT	2015	1962	2070	2072

Note. RT: reading time; Met: metaphoric. Lit: literal.

We found that the reading times were significantly longer for the metaphoric nouns than for the literal nouns and that the subsequent verbs were also read more slowly after a metaphor, reflecting a spill-over effect. However, the endings of metaphoric sentences were read significantly faster than the endings of literal sentences. Table 7 shows the aggregated reading times for the noun and verb, for the endings, and for the whole sentence. As can be seen, the shorter ending times balanced the longer beginning times resulting in no overall difference, which was the result reported by Ortony et al. Our experiment also looked at success in judging the subsequent sentence and found that participants who had read a metaphoric target performed worse, suggesting a comprehension deficit in some cases.

Based on these data, we concluded that participants had only a partial understanding of the metaphors and that, sometimes, they failed to integrate the metaphoric sentences with the preceding context (and thus read the endings faster). An analysis of the endings of target sentences confirmed this conclusion. As seen in Table 6, the endings of the target sentences could be split into two classes: endings related to the passage (e.g., *class* in the targets of the linebacker story from Table 6) and endings that were novel with respect to the passage (e.g., *noisily* for the women story). When looking at reading times for the two classes of endings, we found that participants were faster for the unrelated endings in the metaphoric condition (see Table 8). In the original study, we argued that the unrelated endings offered little help with the process of integration with discourse; therefore, participants may have failed to generate integrations for at least some of the metaphoric sentences with unrelated endings and, thus, may have processed them quickly.

For this experiment, we need an estimate of the similarity between the metaphor or literal and their referent in the text. However, objectively identifying the referent in a text is no easy

Table 8
Ending reading times (ms) from Budiu and Anderson (2002) and model results

	Data		Model	
	Related	Unrelated	Related	Unrelated
Metaphoric noun	811	748	882	799
Literal noun	761	826	869	902

Note. Related stands for reading times of endings related to the context. *Unrelated* stands for reading times of endings unrelated to the context.

task, because relevant features may be diffused across various sentences (for instance, if the word *hens* is used to refer to a group of women that noisily discuss some topic in a meeting, the *female* and the *noisy* features of this referent may be described in several sentences). To solve this problem, we estimate the similarity between the metaphor and its referent by the LSA distance between the metaphoric word and the entire story (up to the target). For consistency (and also because literals were not always identical to their referents—for instance, the literal *athlete* could be used to refer to an arm wrestler), we estimated in the same way the similarity between the literal and its referent. This technique led to relatively low similarity values between literals and their referents. We obtained the results in Table 7 by setting the similarity between the noun and its antecedent in the passage to 0.19 for metaphors and 0.34 for literals. The corresponding LSA similarities for the verbs were 0.23 (metaphors) and 0.44 (literals). These numbers were obtained by averaging the LSA distances between the critical word (metaphoric or literal, noun or verb) and the corresponding passage. These values illustrate that literal meaning is hard to define; as Gibbs (2001) observed, “the very idea of literality carries with it many assumptions about default meaning and processing that are simply unwarranted and not experimentally verified”. Moreover, these low similarities for literals support the idea that there is no dichotomy between metaphors and literals, but, rather, that metaphoric or literal processing is just a matter of degree: the more similar (or salient) the word phrase is to its antecedent, the fewer local failures of comprehension and the smoother the processing.

Next, we discuss the predictions for the aggregated noun-and-verb reading times and for the ending-reading times. The predictions for the sentence-reading times result from summing up the components.

2.2.2.1. Aggregate noun-and-verb reading times. The data in Table 7 indicate that people take longer to read the initial part of a target when it contains a metaphoric noun than when it contains a literal noun. INP predicts these results because it tries hard to find an antecedent for the first word of the sentence. For metaphoric nouns, the chance of finding an antecedent is very low, so the model spends more time in repeated searches for one. On the other hand, for literal nouns, this process needs fewer attempts and is more effective (because the literals are more similar to their referents than the metaphors), so it is faster.

2.2.2.2. Ending-reading times. Consider the situation in the right column of Table 6. For metaphoric sentences with unrelated endings (e.g., *The hens cluck noisily*), most frequently, the given information does not suffice to retrieve any candidate interpretation from the context, because the activation spreading from metaphoric words (e.g., *hens* and *clucked*) is not enough to select any proposition. Therefore, because it was not able to find a candidate interpretation at any point during the processing of the input, at the end of the sentence the model has no hook to perform integration and fails to relate the input to the context; in consequence, the sentence is perceived as isolated. On the other hand, if the ending is related to the context, it can help to find a hook for integration. Indeed, although no interpretation may be found on the metaphoric noun and verb, the presence of an ending related to the context may boost the spreading activation to a level high enough to select a candidate interpretation.

Whenever it has a hook or a final interpretation, INP uses it for integrating the input with the context. Context integration is quite minimalist in INP: it means only updating the semantic representation of the sentence to make sure it is consistent; in a more complex model of discourse processing it may require more elaborate processes. However, the important assumption for simulating this experiment is that integration takes extra time at the end of the sentence. Therefore, INP predicts that, each time when a hook is found, context-integration time adds to the ending-reading times. Hence, unrelated endings of metaphoric sentences should be read faster than related endings, because the latter lead more often to context integration than the former.

An important prediction of INP is that the time to read the noun (be it literal or metaphoric) in this experiment depends on the similarity of the noun to its antecedent in the passage. To test this hypothesis, we did an analysis of the individual nouns to find out how their similarities relate to the performance (in terms of noun reading times and accuracies) on that item. In the original study (Budiu & Anderson, 2002), we had collected ratings of the familiarity of the noun metaphors;[10] we can use these ratings to see if the results are more extreme for the poorer metaphors. First we examined reading times, and found a marginally significant correlation ($r = .32$, $p \approx .1$ for 28 items) between reading times for the noun and rated familiarity (partialing out word frequency). We also wanted to examine whether there was a relationship between rating and success on the subsequent comprehension test. To assess this relationship, we filtered the original 28 items to include only those items where subjects showed a 20% improvement in the literal trials over a baseline performance estimated in the absence of the target sentence.[11] (Many of the questions were answered accurately in absence of the target sentence—16 of 28 with greater than 75% accuracy.) Table 9 shows the seven items that survived this filter (metaphors and our rendition of their antecedents in the story), their rated familiarity, the reading times for the nouns, and the accuracy on the subsequent questions. Generally, there is a relationship in the expected direction between the familiarity rating and the two measures of comprehension success ($r = .89$, $p < .005$ for accuracy, $r = -.63$, $p = .13$ for noun reading times). Interestingly, the LSA values show only a modest nonsignificant correlation with the ratings ($r = -.32$, $p > .400$), in the opposite direction. This failure of LSA to work for individual items suggests that averaging LSAs over several items (as we do in fitting the data of our simulation) may be a more reliable indicator of the average "true" similarity than using individual LSAs to stand for similarities for individual items.

Table 9
Items from Budiu and Anderson (2002) that had a larger than 20% improvement in the literal trials over the no-target experiment

Metaphor/antecedent	Noun RT (ms)	Accuracy	Familiarity
Beast/aggressive supermarket chain	734.93	0.55	1.30
Bud/beautiful girl	566.51	0.95	2.90
Butterfly/graceful ballerina	638.00	0.74	2.20
Cat/man enjoying fire warmth	650.69	0.77	1.70
Duelers/arguing spouses	696.00	0.78	1.70
Iceberg/unresponsive official	671.53	0.75	2.20
Nightingale/soprano	736.85	0.93	2.40

2.3. Moses illusion

The preceding experiments showed how INP (and human participants) can process metaphoric language with varying degrees of success. In the Moses-illusion experiments the same factors that lead to success in the processing of metaphors cause INP to make the same mistakes as people and fail to detect what ought to be glaring contradictions. Erickson and Mattson (1981) were the first who studied Moses (or semantic) illusions. They asked their participants to look for distortions in sentences such as *How many animals of each kind did Moses take on the ark?* Surprisingly, people failed to find the distortions in these questions, in spite of knowing the corresponding undistorted facts (e.g., that Noah, rather than Moses, took the animals on the ark). As a dependent measure in their study, Erickson and Mattson defined the *illusion rate* as the percentage of undetected distortions out of cases in which the correct answer is known. Thus, the illusion rate for a question is based on the number of participants who have the correct knowledge, rather than on the total number of participants.

Reder and Kusbit (1991) followed up on the Erickson and Mattson's (1981) result and introduced a slightly different paradigm, the *gist task*. Unlike the original Erickson and Mattson's task (henceforth called the *literal task*), in which participants had to detect distortions in Moses-illusion type of questions, in the gist task they needed to ignore the distortions and answer the questions as if they were undistorted. For example, the correct answer to the Moses question is *distorted* in the literal task and *two* in the gist task. The gist task is analogous to being asked to comprehend a metaphoric sentence. Whereas for the literal task, the illusion rate is the dependent variable, for the gist task the corresponding measure is the percentage of correct answers.

Table 10 shows the results of Experiment 1 from Reder and Kusbit (1991), comparing latencies for correctly answering distorted questions (e.g., *How many animals of each kind did Moses take on the ark?*) with those for correctly answering undistorted questions (e.g., *How many animals of each kind did Noah take on the ark?*). Whereas in both gist and literal tasks there was no statistically significant difference in latency between the distorted and undistorted questions, participants responded faster in the gist task than in the literal task. Also, in the gist condition, they tended to take longer (but not significantly longer) to answer correctly the distorted questions than to respond to the undistorted questions. These results indicate that in the literal condition people process more carefully the questions than they do in the gist condition.

Although, generally, people find the literal Moses-illusion task difficult, not any distorted question can trick them (e.g., *Who was the first man who walked on the sun*). Ayers et al.

Table 10
Mean response latencies (s) for correct responses in the gist and literal tasks for semantic illusions: data and model

Question	Data		Model	
	Literal	Gist	Literal	Gist
Undistorted	4.25	3.69	4.37	3.48
Distorted	4.29	3.88	4.31	3.84

The data are adapted from Experiment 1 in Reder and Kusbit (1991).

Table 11
Illusion rates in the literal Moses-illusion task and the percentage correct in the gist Moses-illusion task: data and model

Question	Illusion rate (literal)		Error rate (gist)	
	Data	Model	Data	Model
Undistorted	7	10	18	18
Good distortion	46	55	24	21
Bad distortion	29	24	26	31

The data are adapted from Ayers et al. (1996). *Note.* An illusion for an undistorted question is to call it distorted.

(1996) compared illusion rates for good and bad distortions embedded in similar sentences. They looked at three variants of the same question: one containing a good distortion, one containing a bad distortion, and one containing the undistorted term.[12] For example, the three variants could be *How many animals of each kind did Moses take on the ark?* (good distortion), *How many animals of each kind did Adam take on the ark?* (bad distortion), and *How many animals of each kind did Noah take on the ark?* (undistorted term). Ayers et al. conducted an informal rating of the "good" and "bad" distortions and established that the good distortion shared more features with the undistorted term than the bad distortion.[13] Ayer et al.'s results (Table 11) showed that people had most difficulty with the good-distortion questions.

In the simulations for metaphor data we were able to capture the different effects for literals and metaphors by assuming the same cognitive processes; the only difference between the literals and metaphors was that the former were more similar to their antecedents than the latter. In accounting for the Moses-illusion experiments, we assume that the difference between good and bad distortions is also in their degree of "literality" the higher the similarity between the distortion and the undistorted term, the more likely the illusion is to work. (This assumption is confirmed by the ratings conducted by Ayers et al., 1996; van Oostendorp & de Mul, 1990; van Oostendorp & Kok, 1990.) Because the spreading activation reflects semantic similarity, depending on how similar the distorted term in the sentence is to the undistorted term, INP may behave in any of the following ways: (a) it may comprehend the sentence very smoothly, without forming any bugs, and thus not noticing the distortion; (b) it may experience local failures of comprehension (i.e., may form bugs), but still reach to a final interpretation; (c) it may end up with no final interpretation.

In the literal task, INP uses the existence of bugs as a basis for deciding whether a sentence may be distorted. INP searches for interpretations of the sentence among the propositions from background knowledge and answers *undistorted* if it comprehends the probe with no bugs and *distorted* if it has generated any bugs during comprehension. (Remember that bugs correspond to local failures of comprehension—that is, moments when the model did not find any interpretation to match the current words.) In contrast, the model ignores bugs in the gist task, when reaching an interpretation is all that is required. Because INP is more likely to find an interpretation when the similarity between the distorted and undistorted terms is higher, it predicts that (a) in the literal task, the illusion rates for good-distortion questions are higher than those for bad-distortion questions; (b) in the gist task, the percent correct is greater for good-distortion sentences than for bad-distortion sentences. Table 11 presents the illusion rates and percentage correct as resulted from the simulation.[14]

In the gist task, if it has found an interpretation for the current sentence, INP may stop before reaching the end of the sentence and answer according to the current interpretation (provided that it has processed enough words in the sentence). (Participants in Ayers et al., 1996 study also tended to answer before reading the end of the sentence.) The probability of stopping (given that the model has an interpretation) is 0.5 for each new word. Thus, because the chance of finding an interpretation sooner is higher for undistorted sentences than for distorted sentences, on average INP stops sooner for undistorted questions and, thus, tends to answer them faster in the gist task (see Table 10 for the model's response times). This is also why INP captures the difference in latency between the gist and the literal task (i.e., in the gist task, the model sometimes stops before the end). It simulates the participants' tendency in the gist task to give the answer as soon as they have it.

We obtained the results presented in Tables 10 and 11 by assuming that the semantic similarity between the good distortions and undistorted terms was 0.44 and the similarity between the bad distortions and the undistorted terms was 0.33. These values were obtained by averaging the LSA distances between the distortions and the undistorted terms. The LSA values for individual items were modestly correlated with the illusion rates ($r = .33$, $p < .05$). As expected, the illusion rate in the literal task and the percentage of correct answers in the gist task are monotonically increasing functions of the similarity between the distortion and the undistorted term. Table 12 shows how the performance of INP varies for different values of the similarity. The higher the similarity, the higher the illusion rate in the literal task and the accuracy in the gist task. Note that the latency of the correct response in the literal task is not much affected by the similarity; however, in the gist task, distortions of low similarity tend to take longer (because the chance of finding an interpretation for the sentence is smaller).

To summarize, INP accounts for the influence of semantic similarity on the illusion rates: the higher the similarity, the more likely the model is to fall for an illusion or to grasp the gist of the sentence. INP uses the distortion as well as the other information in the sentence as sources of activation; if the distortion is similar enough to the interpretation, then it will still spread enough activation to allow smooth, failure-free comprehension. The difference between the gist task and the literal task is that INP falls for an illusion only if the comprehension was

Table 12
Performance of INP in the Moses-illusion experiments as a function of the similarity between the distorted and the undistorted terms

Task	Measure	Similarity					
		0.18	0.28	0.38	0.48	0.78	1.00
Literal	Illusion rate	0.00	0.07	0.40	0.68	0.92	0.90
	Latency (s)	4.28	4.30	4.32	4.33	4.36	4.38
Gist	Error rate	0.66	0.45	0.25	0.19	0.20	0.22
	Latency (s)	4.61	4.25	3.85	3.66	3.46	3.48

local-failure free in the literal task, whereas in the gist task it disregards memories of local failures and focuses only on the existence of a final interpretation.

2.4. Text priming: Schwanenflugel and White (1991)

So far we have discussed the role of interpretation in integrating the sentence with past knowledge and the current context. However, the interpretation can also facilitate the processing of related words. Priming paradigms are a commonly used method of measuring lexical processing. One task on which priming has been extensively studied is lexical decision. We chose to model an experiment by Schwanenflugel and White (1991) because it was also addressed by Kintsch's (1998) Construction-Integration model, to which we want to make comparisons. Schwanenflugel and White (1991) found that both preceding discourse and local sentence context can influence lexical decision. In their experiment, participants read a short passage followed by an unfinished sentence and then made a lexical decision about a word that ended the sentence. The passage could be consistent, partially consistent, or neutral to the final sentence and the target word could be either locally expected or locally unexpected on the basis of the final sentence fragment. Table 13 shows sample materials used in Experiment 2 in Schwanenflugel and White (1991). Note that, for consistent passages, the locally expected word is consistent with the previous passage, whereas, for the partially consistent passages, the locally unexpected word is consistent with the previous passage (at least with the last complete sentence). For each consistent or partially consistent passage, participants saw a neutral passage (not shown in Table 14) followed by a target that was similar in frequency or length to the target in the nonneutral passage. The neutral context contained four sentences such as *This is the first sentence of this paragraph* and ended with the sentence fragment *The last word of this sentence is*.

Schwanenflugel and White's (1991) results in terms of differences with respect to the neutral passage are presented in Table 13. They show that participants are fastest when the target is expected and the context is consistent. The next fastest case is that of an unexpected target and a partially consistent passage. The 66 ms difference between locally unexpected targets in consistent contexts and neutral contexts was not significantly different from zero in the item analysis. These results indicate that the priming effect depends on whether the target is related

Table 13
Sample materials used by Schwanenflugel and White (1991)

Consistent		*Partially consistent*
The equipment they carried was heavy. They had gotten an early start at dawn. It had been a long day for the guys. The hiking trip was the most strenuous the group had had. The hikers slowly climbed up the		The equipment they carried was heavy. They had gotten an early start at dawn. It had been a long day for the guys. After a treacherous hike, Bill and his friends sluggishly entered their apartment lobby. The hikers slowly climbed up the
	Target words	
Locally expected		mountain
Locally unexpected		stairs

Table 14
Differences in lexical-decision latencies (ms) and error rates (in parentheses) with respect to the corresponding neutral condition in Experiment 2 from Schwanenflugel and White (1991): data and simulation results

Prior context	Target			
	Data		Model	
	Locally expected	Locally unexpected	Locally expected	Locally unexpected
Consistent	81 (0.07)	−66 (0.02)	80 (0.12)	0(0.02)
Partially consistent	42 (0.07)	79 (0.04)	20 (0.04)	70 (0.08)

Note. Positive differences indicate responses that are faster (more accurate) than in the neutral context.

to both the last sentence and to the preceding passage, with the preceding passage having a somewhat larger impact.

In the simulation for this experiment, INP reads the sentences in the passage one by one and, as before, tries to find an interpretation for each of them. The search for an interpretation is among background-knowledge propositions. Each time the model finds an interpretation for a sentence, it also identifies the "script"[15] that contains that interpretation and keeps the script in the focus for a while. Therefore, that script works as a source of activation and, next time when INP searches for an interpretation, favors other propositions from the same script (because these propositions are associated to their script).

Each pair of paragraphs in the study is formed from two scripts: a major script and a minor script. In the example in Table 13, the major script is the hiking trip and the minor script is entering the lobby. All sentences in the consistent paragraph come from the major script, whereas in the partially consistent paragraph the last complete sentence comes from the minor script.

When it starts reading the final sentence fragment (e.g., *The hikers climbed up the*), INP searches for interpretations in the current script (i.e., in the last seen script). If the major script is current, the model will easily find in it an interpretation corresponding to the final sentence fragment. That interpretation is part of the focus in the lexical-decision task and, because it is semantically similar to the locally expected target, it will facilitate its processing. On the other hand, when the minor script is current, INP will often not be able to find an interpretation in that script for the final sentence fragment and might instead switch back to the major script. If it switches to the major script, the lexical decision is facilitated for the locally expected targets; if it keeps the minor script because it was able to find an interpretation, then the locally unexpected target is at advantage. The probability of giving up the minor script depends on the similarity between the final sentence fragment and the minor script. It also depends on the probability of abandoning a script if at some point an interpretation from that script does not match the current word; this probability was set at 0.20. We assumed that the verb (*climbed*) is typical for both scripts and that the similarity between the first word in the final sentence (e.g., *hikers*) and the propositions in the minor script, was 0.13.[16] This latter value should reflect how similar the word is to a generic agent of the script, such as *people*. The specific estimate that we used reflected the LSA similarity between *hikers* and *people*.[17]

Our model of lexical decision is very simple: we assume that, given a string, the model attempts to retrieve a meaning for that string; if it succeeds it responds *yes*, otherwise *no*. The

Table 15
Differences in latency expressed with respect to the neutral context in the simulation as a function of the similarity between the first word of the last sentence (e.g., *hikers*) and the agent of the minor script (e.g., *people*)

Prior context	Target	Similarity				
		0.05	0.23	0.33	0.43	1.00
Consistent	Expected	80	80	80	80	70
	Unexpected	0	0	0	0	10
Partially consistent	Expected	20	20	20	20	10
	Unexpected	60	60	70	70	70

Note. Positive differences indicate responses that are faster than in the neutral context. Latencies are expressed in milliseconds.

success and speed of retrieval depends on the spreading activation from the goal: if the items in the goal (e.g., current interpretation) are related to the word, then the model will respond fast.

Given that we estimated the LSA similarity between the first word of the final sentence (*hikers*) and the agent of the minor script (*people*) based on a single example, we wanted to assess how stable the results of the simulations are. Table 15 shows the predictions of the model (in terms of latency differences with respect to the neutral context) for various similarity settings. (The error rates also vary very little and have values close to the ones in Table 13.) We can see that the model produces virtually the same results independent on the specific similarity value used, for relatively small similarity values (between 0 and 0.43). Naturally, the behavior becomes symmetric for the case when the similarity is 1 (because the last sentence fragment becomes equally consistent with both the minor and the major script).

3. Predictions of INP

In this section we identify some predictions of our model that go beyond the specific simulations in this paper. Perhaps one central prediction of INP is that the ordering of given versus new information within a sentence (or a passage) is a major factor in comprehension. Specifically, the model implies that whenever the given information precedes the new information, an interpretation has higher chances to be found and therefore comprehension is facilitated. We have seen this principle at work in the Gerrig and Healy (1983) and Onishi and Murphy (1993) simulations. However, it can be used in other domains, some discussed in this paper (e.g., Moses illusion[18]). The ordering effect can be extended to the level of multiple sentences or passages. Thus, INP predicts that, when the information relevant to the identification of a script is given in advance, the comprehension of further sentences is easier. This prediction is consistent with studies such as Dixon’s (1987), which showed that texts such as *This will be the picture of a house. Draw a rectangle with a triangle on top* are better understood (in terms of comprehension times and accuracies of execution) than equivalent texts in which the script information was given at the end *Draw a rectangle with a triangle on top. This will be the picture of a house.*

The order effect was also found in memory experiments (e.g., Bransford & Johnson, 1972, who, in their famous "washing clothes" study, showed that participants' memory for a text is improved if they are given the topic in advance). In general, INP can simulate such experiments, assuming that the interpretation found at study, when the sentence is first comprehended, helps situate the sentence in memory and facilitates access to other script propositions that may be retrieved as inferences during recall or recognition (see Anderson et al., 2001; Budiu, 2001 for a more in-depth explanation of these effects).

The domain of text inferences is also one in which INP can make contributions. First, INP predicts that, in normal comprehension, there are no inferences drawn during reading, except for the interpretation of the sentence. The reason for this behavior is the limited time—there are too many syntactic and semantic processes going on to leave time for extra inferencing. However, the interpretation itself can be a source of inferences, if these inferences are needed at a later time. For instance, recalling that a sentence was about folding the laundry from the washing machine may activate other facts about washing clothes, even if those were not studied.

Another domain in which INP makes predictions is the processing of lexically ambiguous words—words with two (or more) meanings. INP predicts that the current candidate interpretation spreads more activation to the related meaning of the ambiguous word than to the unrelated meaning. This effect translates into context having some role in the selection of the meaning, although other factors (e.g., frequency) will also affect the relative activation of the two meanings. Such predictions are consistent with experimental findings (Duffy, Morris, & Rayner, 1986; Rayner & Duffy, 1986; Rayner & Frazier, 1989).

Although we did not elaborate on syntactic processing in this article, one domain where INP may give some interesting insights is the influence of semantic processing on syntax. However, there is one natural extension of INP to this domain—namely, in the case of syntactic ambiguity (be it temporal or final) the candidate interpretation could favor an assignment of thematic roles that is consistent with that interpretation. This behavior would correspond to what the INP does in the case of lexical access, in the Schwanenflugel and White (1991) simulation—there the interpretation favors the lexical access of the incoming words. Similarly, the candidate interpretation has the potential of affecting the syntactic processing, in particular the assignment of thematic roles. Preliminary results of a study run in our laboratory indicate some support for this hypothesis.

4. Conclusions

In his 1988 paper, Kintsch argued that modeling comprehension with a top-down, rule-based system is not a realistic endeavor, because "it is difficult to design a production system powerful enough to yield the right results but flexible enough to work in an environment characterized by almost infinite variability." In this article we showed that a simple production system with a powerful, similarity-based mechanism of spreading activation may offer the right mixture of bottom-up and top-down processes: spreading activation may lead to the selection of the right interpretation and, once that interpretation is selected, it can ensure resilience of comprehension through a flexible matching mechanism.

INP addresses a number of issues in the metaphor literature. First, as other modern models in the field (e.g., Frisson & Pickering, 2001; Giora, 1997; Katz & Ferretti, 2001; Peleg et al., 2001) it postulates the same processes for metaphoric and literal sentences. INP shows that, as Giora (1997) observed, the metaphoric–literal divide, which for so long has fascinated psycholinguists, is in fact a continuum. Since the similarity of the metaphor or of the literal to their referents (which is part of what Giora, 1997 calls "salience" and Haviland & Clark, 1974 call "given" information) drives the processing, literals with low similarity to their context can be processed in the same way as metaphors are and vice versa. Sentence context can increase the amount of "given" information, so the different comprehension patterns for metaphors are associated with different degrees of support from the sentence context. Whereas the importance of the preceding text has been demonstrated (e.g., Inhoff et al., 1984; Ortony et al., 1978), INP emphasizes the different roles that sentence context (be it preceding or following the metaphor) can also play in comprehension. Thus, supportive preceding context can speed up metaphor comprehension as in Gerrig and Healy (1983) and supportive following context can facilitate correct comprehension, although the metaphor may be initially not understood (Budiu & Anderson, 2002; Onishi & Murphy, 1993; see also Budiu and Anderson, 2003). Moreover, lack of supportive sentence context may lead to fast reading but poor comprehension due to lack of integration of the sentence with the larger discourse context (Budiu & Anderson, 2002). Metaphoric sentences typically offer a mixture of "new" and "given" information; comprehension is possible to the extent that the other "given" information in the sentence is enough to relate it to the context. Predicative metaphors do not pose difficulty to INP both because they occur relatively late in the sentence and so are helped by the early part of the sentence and because predicates, be they literal or metaphoric, are typically new information.

INP bears similarity to other process models existent in the field. Perhaps the most prominent sentence-processing theory is Kintsch's (1988) construction-integration model. Kintsch's (1988) model, as INP, is based on associations between words and propositions. More recently, Kintsch addressed the problem of metaphor comprehension in the framework of his CI theory. Specifically, he looked at the comprehension of predicative metaphors of the type A is B. Kintsch used an LSA-based (Landauer & Dumais, 1997) knowledge representation, in which the strengths of the association-network connections are given by their LSA distance. To apply the CI theory to this knowledge structure, Kintsch (2000) defines a predication algorithm for understanding A-is-B metaphors. The predication involves integration in a network formed by A, B and relatively close neighbors of B. The meaning of the metaphor is the centroid between A, B and the most active terms in the network, after the system stabilizes. Kintsch argues that the same theory can be applied to the understanding of literal predications.

In comparing Kintsch's metaphor-comprehension theory with INP, we must keep in mind that there are important scope differences between the two. Specifically, Kintsch's theory only addresses A is B predicative metaphors, whereas our model addresses metaphoric sentences that can be predicative or anaphoric.

Kintsch evaluates his theory in terms of three empirical results that it captures: (1) metaphors are not reversible, (2) activating the literal meaning of the metaphor can harm the comprehension of the metaphor, (3) understanding metaphors is similar to understanding lexically-ambiguous words. Next we discuss how INP fares on the same tests.

The irreversibility of metaphors refers to the difference between A is B and B is A—compare, for instance, *Her surgeon is a butcher* with *Her butcher is a surgeon*, or *His marriage is an icebox* with *His icebox is a marriage*. Due to the asymmetry of Kintsch's predication algorithm, his theory agrees with the data. INP also would find different interpretations for each sentence in such a pair. To see why, let us assume that a proposition such as *Her butcher is a surgeon* was initially part of the model's prior knowledge.[19] Then that proposition would be the interpretation assigned by the model to the sentence *Her butcher is a surgeon*, but consider what proposition would be chosen as an interpretation for *Her surgeon is a butcher*. In processing the end of the sentence *Her surgeon is a butcher*, if INP has not found a valid interpretation yet, it will look for a proposition in which *butcher* is part of the predicate. It could retrieve *Her butcher is a surgeon*, but that proposition would be rejected because *butcher* would fail the matching test with the corresponding concept (i.e., *surgeon*) in the predicate role of the interpretation.[20] After rejecting this proposition, INP might retry and find an interpretation such as *Her surgeon is rough* and might well accept this if *rough* and *butcher* are similar enough.

With regard to the literal meaning interfering with metaphor understanding, Kintsch suggests that preceding a metaphoric sentence such as *My lawyer is a shark* with a literal sentence such as *Sharks can swim* leads to people taking longer to understand the metaphor. In Kintsch's predication model, this effect can be simulated by asserting that *Sharks can swim* activates those neighbors of the predicate that are related to the literal meaning of *shark*; they start with some positive activation at the beginning of the integration and it takes longer for this prior activation to be washed out. In INP, if the interpretation or script associated to the previous sentence (*Sharks can swim*) is still in focus (as in the text-priming simulation), then that interpretation could interfere with the retrieval of a correct interpretation for *My lawyer is a shark* and thus delay the comprehension.

Another phenomenon cited by Kintsch as predicted by his theory is that metaphor understanding is similar to lexical-ambiguity resolution. Both Kintsch's model and ours predict that in the presence of a supporting preceding context, the right meaning of a homonym is retrieved. For homonyms, if the current interpretation (promoted by preceding context) favors a certain meaning of the homonym, that meaning will be retrieved due to extra activation from the interpretation. This behavior is similar to the text-priming simulation, where we showed that the current interpretation can facilitate the processing of words related to it.

Another feature of Kintsch's model is that it proposes the same comprehension process for both literal and metaphoric predication. This is also true of INP—if the metaphor occurs at the end of the sentence (and it does for predicative metaphors) and if the model has found the correct interpretation, it will integrate the metaphor smoothly into the interpretation, in the same way it would do with a literal sentence. If the correct interpretation was not found (i.e., the context was scarce), then the search for an interpretation occurs for both the metaphor and the literal. The only potential for a difference between metaphors and literals is their different similarity to their referents.

Although INP can understand sentences embedded in context, it is still a rudimentary model of text processing. For instance, it does not deal with binding or pronoun resolution. However, it is interesting to compare it with existent text-processing theories. One current view, popular among several researchers (Albrecht & Myers, 1998; Cook, Halleran, & O'Brien, 1998; Myers & O'Brien, 1998; Noordman & Vonk, 1998; Sanford & Garrod, 1998) is the memory-based–text-processing (MBP) approach. The central assumption of MBP is that the processes involved during text comprehension are an effect of more basic memory processes: reading new information evokes older information (from the text) through an activation-spreading process and, thus, makes that information readily available. (This process is called resonance.) Whatever inferences are made during reading, they are not explicit, but rather due to the old information being "dumbly" activated. INP is consistent with this view, although the information evoked from prior text is limited to one single proposition that is "dumbly activated" (through the process of spreading activation) by the elements in the current sentence. In both MBP and INP, this old information is at the core of text inferences.

INP has a syntactic component which has been only briefly discussed in this paper. It is still somewhat rudimentary (it does not cover a lot of English constructions—for instance, relative clauses) and does not yet deal with syntactically ambiguous sentences. However, as discussed in the section on predictions, the resolution of syntactic ambiguity is a domain where INP may be able to show its predictive power. Specifically, the interpretation may be a factor in resolving syntactic ambiguity. In this respect, INP is relatively similar with constraint-satisfaction models (e.g., MacDonald, 1997; MacDonald, Pearlmutter, & Seidenberg, 1994), which assume that semantic factors such as the lexical properties of the words involved, the context, and the frequency of the structure all influence the syntactic-ambiguity resolution. Indeed, INP selects its candidate interpretation as the most active proposition in memory that matches the sentence context. A proposition that corresponds to a frequent structure will presumably be more active than another proposition; moreover, such a proposition needs (by definition, since it is the result of the storing of language input) to satisfy the lexical constraints imposed by the words involved.

One question that can be asked about INP is to what extent its predictions reflect just LSA similarity, given that the model uses similarity values that are taken from the LSA theory. Whereas LSA is a theory of similarity, it does not offer a process explanation for language processing. As such, it cannot distinguish, for instance, between situations like those in Gerrig and Healy's (1983) experiment, where only the order of the words in the sentence is manipulated. It also cannot explain the time course of processing in metaphor understanding studies such as Budiu and Anderson (2002), or the difference between the anaphoric and predicative metaphors as in Onishi and Murphy (1993). However, to convince ourselves, we carried out a small experiment. We used practice items from the Accuplacer reading comprehension test;[21] these items were short texts, followed by four multiple choice questions. We used INP to parse a modified, syntactically simpler version of the texts and to say which of the four choices is true, based on the text; we also used LSA to compute the distance between the four choices and the text and assumed that the answer was the one with the highest LSA. Our model achieved an accuracy of 50%, whereas LSA was lower than chance (15%).

Another issue that can be raised about INP is how scalable it is. Inherently, the background knowledge with which INP operates is limited to a small number of propositions; however, people probably operate with lots of facts in their background knowledge. Budiu (2001) presents a simulation of the semantic part of INP when the number of propositions in the background knowledge is relatively large. We were especially interested in two results: (1) whether, presented with a sentence, the model finds the correct interpretation for it in the database, and (2) whether the time to find this interpretation is reasonable. The database we used was obtained by running a query for noun–active verb–noun sentences on the Brown corpus (http://www.ldc.upenn.edu/ldc/online/). The Brown corpus was compiled in the early 1960s at Brown University under the direction of W. Nelson Francis and Henry Kucera. It contains 500 text samples of circa 2000 words each, representing 14 categories(e.g., literature, fiction, government). The result of the search consisted of 457 propositions that satisfied the pattern agent–verb–object. We chose such 3-concept items mainly because such sentences are simple enough syntactically. We presented random sentences from the database to the model and looked at the final interpretation. The results were encouraging: for a range of parameters, the simulation indicated that the model was able to find the correct interpretation in 95% of cases. Moreover, it did so reasonably fast (the average number of candidate interpretations that were examined by the model before finding the correct one was less than 0.7).

Finally, we would like to close this paper by reviewing what we think are the important aspects of our theory of semantic sentence processing. One point that has been in the background throughout the paper, but we would like to stress here is that the theory provides a complete computational model that goes from the parsing of words to the interpretation of a sentence. It has strong commitments to the real-time processing of sentences. These commitments come from the fact that it is embedded in the ACT-R architecture. That architecture places significant constraints on how much information can be processed in any particular time. Although the theory allows for massively parallel subsymbolic activation computations, these computations must conclude in the selection of a single production to fire and each cycle of production firing must take at least 50 ms. This constraint forces a minimalist style of information processing in which the model counts on the parallel activation to come up with an interpretation and only minimally checks that particular interpretation. The fact that we can successfully process sentences at a speed that humans do and with the behavioral profiles that humans display is a non-trivial accomplishment. The successful embedding of INP in the ACT-R architecture has a number of consequences. Of course, it conveys some credit to the architecture. However, more importantly, it says that sentence processing is not unlike the other sorts of information-processing tasks that ACT-R has modeled.

Over and above its general computational embodiment, INP has made a number of commitments as to the nature of semantic sentence processing. These commitments were driven by the empirical results that we reviewed. Probably, the most striking feature of the model is the claim that sentence processing involves an obligatory search for an interpretation and that this interpretation starts with the first phrase and continues incrementally throughout the sentence. This interpretation of the sentence allows one to recognize what one already knows and relate new information to that knowledge. It also speeds up lexical processing and other low-level aspects of sentence processing. In INP the interpretation is the linchpin of the comprehension process. The incremental nature of interpretation processing has been essential in our accounts of many of the phenomena in the paper.

The other striking commitment has been to similarity-driven processing. The associations that fed the spreading activation processes were based on similarities and, in general, we found Landauer and Dumais's (1997) LSA a useful vehicle for assessing similarities. This commitment has been essential also in our accounts of many of the behavioral phenomena. Moreover, it has the implication that literality is a matter of degree and that metaphor and semantic illusions are just relatively extreme examples of the kind of processing that is essential to our interpretation of every sentence.

Notes

1. Tel.: +1-412-268-2788; fax: +1-412-268-2844.
2. This 50-ms-per-step assumption has been supported in many domains and in many production-rule architectures beside ACT-R (Just & Carpenter, 1992; Meyer & Kieras, 1997; Newell, 1990) and proves a defining constraint in the ACT-R models.
3. The syntactic representation is inspired by the X-bar theory (Jackendoff, 1977) and is described in more detail in Appendix C.
4. Note that the activation spread from a source does not depend on how active that source itself is.
5. That the process of finding an interpretation is not invoked on every word does not contradict the incrementality of language—the syntactic component of INP, which builds both syntactic and semantic representations, does act on each word and takes decisions even when prepositions or auxiliaries are encountered.
6. The similarity between various meanings is given as input to the model, as discussed in the section on semantic similarities.
7. Technically, the activation of the propositional link involving the concept is raised above the retrieval threshold.
8. As in Anderson and Lebiere (1998), there is a large correlation (0.936) between the retrieval thresholds and latency factors. This reflects the fact that latency is relatively constant at threshold. Thus, in a sense, the only free parameter is the retrieval threshold.
9. The target sentence contained a noun, a verb and an ending; the noun and the verb could be either literal or metaphoric. In this discussion we aggregate our data over the two verb types.
10. Chiappe and Kennedy (2001) looked at similarity as a measure of the number of features that the two components of the metaphor had in common and found a highly positive, highly significant correlation—$r = .84$—between familiarity and similarity.
11. The baseline performance represented the performance in Experiment 2 in Budiu and Anderson (2002). In that experiment, participants read the same story without the target sentence, and then answered the same comprehension question.
12. The authors used their intuition to design the good and bad distortions and, then, confirmed their choices by the rating described below.
13. This informal rating agrees with the results of Oostendorp and de Mul and co-workers (van Oostendorp & de Mul, 1990; van Oostendorp & Kok, 1990), who conducted a more rigorous rating.
14. Note that participants respond *distorted* even to undistorted questions. We simulate this bias by making the model answer *distorted* in 10% (respectively, 20%) of the cases when it did not detect any distortion in the literal task (respectively, in the gist task). The different biases in the literal and gist task reflect the fact that, whereas in the literal case, the illusion rates in the undistorted case include only "distorted" answers, the error rates in the gist task also include wrong answers. Refer to Reder and Kusbit (1991) for a more detailed analysis of the types of error in Moses-illusion tasks.

15. We use the term "script" as meaning a set of propositions corresponding to the same prototypical situation (e.g., eating in a restaurant). This script need not be the structure defined by Schank and Abelson (1977).
16. The similarities between *climbed* and the verb in the corresponding script (be it minor or major) proposition was set to 1, and so was the similarity between *hikers* and the agent of the corresponding proposition in the major script.
17. Unlike for the other simulations, we did not have access to all the materials used in the Schwanenflugel and White's (1991) study. Even if we had, we would also need data on agents of the minor scripts.
18. For Moses illusion, there is some evidence that moving the distortion in focus using cleft sentences may actually decrease the illusion rate (Bredart & Modolo, 1988); it is possible that this effect be a position effect, since focused nouns in cleft sentences occur at the beginning of the sentence (e.g., *It was Moses who took two animals of each kind on the ark*. Bredart and Docquier (1989) also showed that this effect holds when capitalization of the distorted term is used; however, Kamas, Reder, and Ayers (1996) replicated the study and used a bias–sensitivity analysis to prove that capitalization mainly affected participants' bias towards calling a sentence distorted rather than their sensitivity to distortions. Jaarsveld, Dijkstra, and Hermans (1997) investigated position effects on Moses illusion, but, since they did not include an undistorted condition in their experiments, it remains to be shown that their result was indeed an effect of the manipulation.
19. This assumption is unrealistic, but it represents the most unfavorable case for our model; in that situation INP would be most likely to end up with the same interpretation for both sentences.
20. Note that *butcher* in the sentence is not matched against *butcher* in the interpretation because they do not share the same role.
21. Accuplacer is a placement test used in community colleges. We preferred it over other tests available (e.g., Toefl) because the texts and the questions were shorter and somewhat simpler.
22. Note that the switch time does not depend on the position of the word on which it occurs.

Acknowledgments

This research was supported by National Science Foundation Grant BCS997-5-220. We thank Jaime Carbonell, David Plaut, and Lynne Reder for their input on the development of the theory and its application to various data sets.

Appendix A. An overview of ACT-R

ACT-R assumes that human knowledge is structured in two categories: declarative and procedural. The declarative knowledge refers to facts such as *Stockholm is the capital of Sweden* or *2* + *2* = *4*; in ACT-R, these facts would be represented as *chunks*. The procedural knowledge corresponds to knowledge about carrying out actions (for instance, about performing addition or about driving) and is expressed in ACT-R in the form of *productions*. ACT-R claims that human cognition occurs as the result of the interaction between procedural and declarative knowledge.

A.1. Chunks

ACT-R chunks encode "small, independent patterns of information" (Anderson & Lebiere, 1998). The slots structure the information within a chunk. For example, we could represent two chunks encoding that Stockholm is the capital of Sweden and that Oslo is the capital of Norway in the following way:

Sweden-fact	*Norway-fact*
Is a capital-fact	Is a capital-fact
Country *Sweden*	Country *Norway*
City *Stockholm*	City *Oslo*

Chunks are characterized by their *activation*, which is a quantity reflecting how often and how recently the chunk was used in the past and how relevant it is to the current context. Activation plays an important role in the retrieval of the chunk. We talk later in this section about how chunk activation is computed.

A.2. Productions

A production is an if–then rule with a *condition* side, containing one or more conditions, and an *action* side, specifying a number of actions. If the conditions in the condition side are fulfilled, the production can be *fired* and the actions in the action side can be executed. The following example is the verbal description of an ACT-R production for answering a question about the capital of a country:

IF	the goal is to say the capital of the country *c* and *x* can be retrieved as the capital of country *c*
THEN	say *x*

The first condition of any production always refers to the current *goal*; the other conditions are typically *memory retrievals*. The goal denotes a chunk that corresponds to the current focus of attention in the system. The other condition type allowed in the condition side of a production is a memory retrieval: it indicates that a chunk must be retrieved from memory and that it must match the pattern specified in the production.

Cognition in ACT-R emerges from a set of productions that fire in some constrained order; each production can retrieve information from memory and use it to modify the current goal. A production takes at least 50 ms to fire; a production that performs time-consuming actions such as key presses or memory retrievals can take longer. Moreover, two productions cannot fire in parallel. Hence, the more productions that fire to perform a task, the longer the overall time to complete that task. This observation has an important implication for an ACT-R model of sentence processing: if the goal is to match human comprehension speed, the model's complexity (in terms of the total number of productions fired) must be relatively low.

A.3. Subsymbolic computation

The chunks and the productions form the symbolic level in ACT-R. Subsymbolic, continuous quantities govern which production should be fired next or which chunk should be retrieved and how long its retrieval takes. Unlike the symbolic computation, which is serial, multiple computations of these subsymbolic quantities are carried out in parallel.

Chunk activation is one such quantity, which controls the retrieval of a chunk. It reflects how frequently and how recently the chunk was used and it also depends on the context in which the chunk occurs. The context is formed by the chunks in the goal: each of them spreads activation to other chunks. Thus, the activation A_i of a chunk i is the sum of a *base-level activation* and a *spreading activation*, as specified by the following equation:

$$A_i = B_i + \sum_j W_j S_{ji} \tag{A.1}$$

The summation is done over all goal slots j. B_i is the base-level activation of chunk i and depends on the usage history of this chunk. In all simulations in this study we use a constant value for B_i, arbitrarily set to 0 and corresponding to the assumption that most chunks used in the tasks are well known and their activation does not vary over the course of the experiment. The other component of the activation is the spreading activation: each element j that is part of the current focus of attention (i.e., of the current goal) spreads an amount of activation to the chunk i proportional to the association S_{ji} between chunks j and i. W_j reflects the splitting of "attention" among the elements in the focus. By default, ACT-R sets $W_j = 1/n$, where n is the number of elements in focus. In INP, the associations S_{ji} reflect semantic similarities among chunks. In Appendix B we discuss how these similarities are computed.

A chunk can be retrieved only if its activation is greater than a fixed retrieval threshold, τ. ACT-R activations are noisy: a random value is added to the magnitude computed by Eq. (A.1). The noise comes from a logistic distribution with variance σ. As a consequence, a chunk j is retrieved with a probability given by the following equation:

$$P_j = \frac{e^{A_j/t}}{\sum_i e^{A_i/t}} \tag{A.2}$$

where P_j is the probability of chunk j being retrieved. t is a constant dependent on the noise variance σ: $t = \sigma\sqrt{6}/\pi$. In all our simulations we use the value $t = 0.35$. The summation in the denominator is done over all chunks that match the retrieval condition and also includes a term corresponding to the retrieval threshold τ. According to Eq. (A.2), how likely a chunk is to be retrieved depends on how much larger its activation is, compared with the activations of other chunks and with the retrieval threshold.

The activation A_j of chunk j also influences the time to retrieve that chunk: the higher the activation, the faster the retrieval. The relationship between activation and retrieval latency is described by the following equation:

$$T_j = F\,\mathrm{e}^{-A_j} \tag{A.3}$$

where F is a constant latency factor.

Appendix B. Semantic similarities

B.1. Semantic similarity and associative strength

Given the similarity $\sigma(j, i)$ between two chunks i and j, we calculate the associative strength S_{ji} between them using a linear function of the similarity:

$$S_{ij} = C + M \times \sigma(j, i) \tag{B.1}$$

where C is a base associative strength and M is a positive multiplier. C is a negative quantity, indicating that two items can be positively associated only if they are similar enough. In all our simulations, we use $C = -16$, $M = 21$ if one of the chunks is a proposition and $M = 32$ otherwise. Thus, most of the activation spread from an item is negative.

B.2. Calculation of semantic similarities for composite structures

INP takes as input the similarities between words and then, based on them, it computes similarities involving more complex structures such as composite meanings (e.g., *college students*) and propositions. In the simulations in this article we use LSA (Landauer & Dumais, 1997) to set the basic similarities for words. Next we discuss the rules used by the model for deriving composite similarities, using these basic similarities. Note that we assume (as does LSA) that the basic similarities between words are symmetric and all of our composite similarities are defined so that they, too, are symmetric.

B.2.1. Similarity between two meanings

Suppose c_1 and c_2 are two meanings. Let the concepts $c_{11} \ldots c_{1n}, c_{21} \ldots c_{2m}$ be children of c_1 and c_2, respectively. (If either c_1 or c_2 is atomic, it can be regarded as a composite meaning with a unique child—the atomic meaning itself.) Then the similarity between the two meanings is defined as follows:

$$\sigma(c_1, c_2) = \frac{1}{k} \sum_{i=1}^{n} \sum_{j=1}^{m} \sigma(c_{1i}, c_{2j})$$

where k is the maximum between n, the number of children of c_1 and m, the number of children of c_2. One can prove that this definition is symmetric.

Let us go through an example to see how these similarities are calculated. Suppose c_1 and c_2 are *college students* and *preschool students*. Then c_1 has two children, c_{11}—*college* and c_{12}—*students*; similarly, the children of c_2 are *preschool* and *students*. In our case, both c_1 and c_2 have two children each, so $m = n = k = 2$. Each of the children is atomic, so the similarities between them are given as input to the model. Suppose that the similarity between *college* and *preschool* is 0.08, the similarity between *college* and *students* is 0.5, and the similarity between *preschool* and *students* 0.13. Then the similarity between the two meanings c_1 and c_2 is $1/2(\sigma(c_{11}, c_{21})+\sigma(c_{11}, c_{22})+\sigma(c_{12}, c_{21})+\sigma(c_{12}, c_{22})) = 1/2(0.08+0.5+0.13+1) = 0.86$.

B.2.2. Similarity between a meaning and a proposition

Suppose the concepts $c_1 \ldots c_n$ occur in the proposition p. Then the similarity between a meaning m and the proposition p made of those concepts is defined as follows:

$$\sigma(m, p) = \sum_{i=1}^{n} \sigma(m, c_i), \qquad \sigma(p, m) = \sum_{i=1}^{n} \sigma(c_i, m)$$

where $\sigma(m, c_i)$ is the similarity between the meaning m and the concept c_i and $\sigma(c_i, m)$ is the similarity between c_i and m.

For the proposition *Professors teach students*, $n = 3$ and the similarity between the meaning *students* and that proposition is the sum of similarities between *students* and each of the meanings *professors*, *teach*, and *students*, which is $0.5 + 0.5 + 1 = 2$, provided that the similarity between *students* and *professors* or *teach* is 0.5 and the similarity of *students* to itself is 1.

B.2.3. Similarity between two propositions

Assume that proposition p_1 is made of concepts $c_{11} \ldots c_{1k}$ and proposition p_2 is made of concepts $c_{21} \ldots c_{2l}$ and, moreover, that c_{1i} and c_{2i} have the same thematic role (i.e., agent, patient, etc.), for all i. Then the similarity $\sigma(p_1, p_2)$ between p_1 and p_2 is defined as follows:

$$\sigma(p_1, p_2) = \frac{1}{n} \sum_{i=1}^{m} \sigma(c_{1i}, c_{2i})$$

where $\sigma(c_{1i}, c_{2i})$ is the similarity between concepts c_{1i} and c_{2i}, m is the number of common roles in p_1 and p_2, and n is the number of roles that occur in at least one of p_1 or p_2.

If the two propositions were *Professors teach students at college* and *Parents protect children*, then m is 3 (agent, patient, and verb are roles in both propositions), n is 4 (because the roles occurring in at least one of the two propositions are agent, patient, verb, and place oblique) and c_{11} is *professors*, c_{21} is *parents*, c_{12} is *teach*, c_{22} is *protect*, and c_{13} is *students*, c_{23} is *children*.

The similarity between the two propositions is $1/4(\sigma(c_{11}, c_{21})+\sigma(c_{12}, c_{22})+\sigma(c_{13}, c_{23})) = 1/4(0.08 + 0.11 + 0.12) = 0.08$, where the mutual similarities between the agents, verbs, and patients in the two propositions are 0.08, 0.11 and 0.12, respectively.

B.2.4. Similarity between a meaning or proposition and a semantic link

This is just the similarity between the meaning or proposition and the meaning in the *child* slot of the link (see Fig. 1 for an example of semantic link). Also, the similarity between the semantic link and the meaning or proposition is defined as the similarity between the *child* of the link and the meaning or proposition.

Appendix C. Syntactic representation and the syntactic processor in INP

In this Appendix, we discuss in more detail the syntactic representation and the functions of the syntactic processor.

C.1. Syntactic representation

The syntactic representation is inspired by the X-bar theory (Jackendoff, 1977); however, it is somewhat simplified (some X-bar assumptions are violated). Thus, for the sentence *The college students were taught by lecturers*, the syntactic representation corresponds to a parse tree of the sentence (see Fig. 2, last syntactic tree on the page): the leaf nodes are words (e.g., *the*, *students*, *good*) and the interior nodes are nonterminals (e.g., $NP1$, $VP1$, $N'1$). Both nodes and edges of the tree are represented as ACT-R chunks, with edge chunks bearing all the structural information in the tree. Thus, the chunk that corresponds to the edge between $N'2$ and the word *college* points to the parent chunk (i.e., to $N'2$), to the child chunk (i.e., to *college*), and to the sentence (in a slot called *context*). It also contains role information (i.e., that the child is a head of its parent) and information about the parent's type of nonterminal (i.e., N').

C.2. Syntactic component of INP and the interaction between the syntactic and semantic processors

The syntactic component is based on the model described by Anderson et al. (2001), but is somewhat more complex, being able to deal with sentences that contain more elaborate noun phrases (e.g., *college students*, *professors of good reputation*, *athletes from each country*) and also with interrogative structures (e.g., *How many people did the driver take on the bus*). The syntactic component processes each new word in the sentence as it is "read" and builds a syntactic and a semantic representation for the input sentence.

Next, we discuss in more detail how INP processes the sentence *The college students were taught by lecturers* and how it generates the representations in Fig. 2 (still more detail can be obtained by running the simulations available on the web, at http://www.act-r.psy.cmu.edu). First, the model reads the word *the*; it retrieves the meaning and the category of this word (i.e., determiner) and it starts building the sentence representation. Because *the* is a noun determiner, INP knows that it is part of a noun phrase, so it builds the node $NP1$ and it links it to *the*. The model also creates the node *Sent* corresponding to the current sentence and connects it with the noun phrase $NP1$. For the semantic representation, INP creates a node for the proposition corresponding to the input sentence and links it to a noun-phrase meaning. At this point the model does not know what this latter meaning will be, nor does it know what the thematic role (e.g., agent, patient) of the noun phrase will be. Later, when the word *college* is read, INP updates both its syntactic and semantic representations to include it: the node *college* is parsed as the head of the noun phrase introduced by *the* and also as the head of the meaning of that noun phrase. At this point, INP does not consider the noun phrase complete (it assumes that further arguments may come), so it postpones the search for an interpretation for the current input sentence. Next, the word *students* is read and the model modifies the noun phrase $N'1$ to have the head *students* and creates a new node $N'2$ to be the parent of *college* and an argument of $N'1$. The semantic representation is also modified to encode that the noun-phrase meaning is a composite meaning formed by the meanings *students* (as head) and *college* as argument. At this point, the semantic part of INP comes into play: because it has accumulated two noun heads (for $N'1$ and $N'2$) (and the likelihood to be at the end of a noun phrase has therefore increased), the model attempts to guess the meaning of the input sentence based on the words read. Thus, it looks for a fact in the background knowledge that involves the composite meaning *Meaning1* (i.e., *college students*) and makes that fact its current candidate interpretation (e.g., *College students live in dorms*). When the next word *were* comes in, INP updates the parse tree, but, as before, it waits to complete its verb phrase before validating its semantic interpretation. After the main verb *taught* is read, the model knows that the sentence is in passive voice and updates *Meaning1* (i.e., *college students*) to be a patient in the semantic representation. It also augments the parse tree and the semantic tree with the verb information. Then the semantic model, invoked on the verb, checks whether the current word phrase matches the verb in the current candidate interpretation; in our case it does not, so the model needs to search for another interpretation involving *Meaning1* and *taught* and in which *taught* is a verb. Let us assume that the model selects *Professors teach college students* and this fact becomes our current candidate interpretation. The process of building the semantic and syntactic trees continues in the same way for the last words of the sentence; the semantic model is again invoked on the word *lecturers*: if the meanings of *lecturers* and *professors* are similar enough, *lecturers* will match the agent of the current candidate interpretation (*Professors teach college students*), which will become the final sentence interpretation.

Appendix D. Independence on background knowledge structures

The predictions in the Gerrig and Healy (1983) simulation are independent on the particular configuration of the background knowledge. Here we show that the number of interpretation switches, which is the principal component of the comprehension time, is higher for metaphor-first sentences than for metaphor-last sentences. We prove this assertion only for the case of three-concept sentences.

Let f be the probability of finding the right interpretation (*The stars filled the night sky*) for the metaphor-last sentences after reading the first concept (*The night sky*) and let s be the probability of finding the right interpretation on the second concept (i.e., after reading both *The night sky* and *was filled*). Let us also assume a certain probability r of rejecting a wrong interpretation. Moreover, suppose that the model never searches for more than one candidate interpretation per input concept. We are interested in estimating the expected number of interpretation switches for metaphor-last sentences. A switch can happen on the second concept (*filled*), or on the third concept (*drops of molten silver*), or on both. The probability of having only one switch on the second concept is $(1-f)rs+(1-f)r(1-s)(1-r)$: this sum corresponds to the case when a wrong interpretation is selected on the first concept, then it is rejected on the second and replaced either with the right one (first term) or with a wrong interpretation, which fails to be rejected on the third concept. The probability of having only one switch on the third concept is $(1-f)(1-r)r$ (we assume that, given all three concepts, the probability of finding the right interpretation is 1). The probability of switching on both the second and the third concepts is $(1-f)r(1-s)r$. Thus, the expected number of switches performed for a metaphor-last target is:

$$\begin{aligned} N_2 &= (1-f)rs + (1-f)r(1-s)(1-r) + (1-f)(1-r)r + 2(1-f)r(1-s)r \\ &= (1-f)r(2-rs) \end{aligned}$$

For metaphor-first sentences, there is also a possibility of a switch on the second concept, if the interpretation selected on the first does not match it, or on the third concept, or on both. We can assume that the chance of selecting the right interpretation on the first concept is 0, as is the chance of selecting it on the second concept (i.e., you cannot guess a star interpretation after reading *Drops of molten silver* and *filled*). Then, if r is, as before, the probability of rejecting a wrong interpretation, there is a $r(1-r)$ chance of having a switch on the second concept only (that would mean that a wrong interpretation would be final). The probability of having a switch only on the third word is $(1-r)r$ and the probability of switching twice is r^2. Therefore, the expected number of switches for metaphor-first sentences is:

$$N_1 = r(1-r) + (1-r)r + 2r^2 = 2r$$

Then, we compute the difference in the number of switches between the two conditions:

$$N_1 - N_2 = 2r - (1-f)r(2-rs) = r(2f + rs(1-f)) \geq 0, \quad \text{because } 0 \leq f, r, s \leq 1$$

Therefore $N_1 > N_2$. We have shown that the expected number of switches is higher for metaphor-first sentences than for metaphor-last sentences and, therefore, the model takes longer to process the former.[22]

This demonstration is actually pessimistic, because it assumes equal cost of switches for metaphor-first and metaphor-last targets. In fact, for the latter, switches take less time because of interpretation priming: the old candidate interpretation (which was just rejected) helps the selection of related interpretations. For metaphor-last sentences, although initial interpretations may be wrong (e.g., after reading *The night sky was filled*, a possible candidate proposition is *The night sky was filled with airplanes*), in most cases they are more related to the correct interpretation than the bad candidates for metaphor-first sentences. For instance, suppose that the candidate interpretation after reading *Drops of molten silver filled* is *Drops of molten silver filled the bowl*; the bowl interpretation and the correct stars interpretation (*The night sky was filled with stars*) are less similar than the airplane interpretation and the stars interpretation. Thus, less activation spreads from the goal in the case of metaphor-first sentences and the interpretation switch is more expensive than for metaphor-last sentences.

References

Albrecht, J., & Myers, J. (1998). Accessing distant text information during reading: Effects of contextual cues. *Discourse Processes, 26,* 87–107.

Anderson, J., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38,* 341–380.

Anderson, J., Budiu, R., & Reder, L. (2001). A theory of sentence memory as part of a general theory of memory. *Journal of Memory and Language, 45,* 337–367.

Anderson, J., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.

Ayers, M., Reder, L., & Anderson, J. (1996). *Accepting false information now and believing it later: Partial matching and false information in the Moses illusion.* In preparation.

Barton, S., & Sanford, A. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory and Cognition, 21,* 477–487.

Bransford, J., & Johnson, M. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717–726.

Bredart, S., & Docquier, M. (1989). The Moses illusion: A follow-up on the focalization effect. *Cahiers de Psychologie Cognitive Current Psychology of Cognition, 9,* 357–362.

Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica, 67,* 135–144.

Budiu, R. (2001). *The role of background knowledge in sentence processing*. Unpublished doctoral dissertation. Pittsburgh, PA: School of Computer Science, Carnegie Mellon University (Available as Technical Report No. CMU-CS-01-148).

Budiu, R., & Anderson, J.R. (2000). Integration of background knowledge in sentence processing: A unified theory of metaphor understanding, semantic illusions and text memory. In: N. Faatgen & J. Aasman (Eds.), *Proceedings of the 3rd International Conference on Cognitive Modelling* (pp. 50–57). Netherlands: Universal Press.

Budiu, R., & Anderson, J. (2001). *Word learning in context: Metaphors and neologisms.* Tech. Rep. No. CMU-CS-01-147. Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.

Budiu, R., & Anderson, J. (2002). Comprehending anaphoric metaphors. *Memory and Cognition, 30,* 158–165.

Budiu, R., & Anderson, J. (2003). Verification of sentences containing anaphoric metaphors. In *Proceedings of the 5th international conference on cognitive modeling*. Bamberg, Germany.

Chiappe, D., & Kennedy, J. M. (2001). Literal bases for metaphor and simile. *Metaphor and Symbol, 16,* 249–276.

Cook, A., Halleran, J., & O'Brien, E. (1998). What is readily available during reading? A memory-based view of text processing. *Discourse Processes, 26,* 109–129.

Dixon, P. (1987). The processing of organizational and component step information in written directions. *Journal of Memory and Language, 26,* 24–35.

Duffy, S., Morris, R., & Rayner, K. (1988). Lexical ambiguities and fixation times in reading. *Journal of Memory and Language, 27,* 429–446.

Erickson, T., & Mattson, M. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20,* 540–552.

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

Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language, 37,* 331–335.

Gerrig, R., & Healy, A. (1983). Dual processes in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology Memory and Cognition, 9,* 667–675.

Gibbs, J. R. W. (1992). Figurative thought and figurative language. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.

Gibbs, R. (1990). Comprehending figurative referential descriptions. *Journal of Experimental Psychology: Learning Memory and Cognition, 16,* 56–66.

Gibbs, R. (2001). Evaluating contemporary models of figurative language understanding. *Metaphor and Symbol, 16,* 317–333.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics, 8,* 183–206.

Haviland, S., & Clark, H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512–521.

Inhoff, A., Lima, S., & Carroll, P. (1984). Contextual effects on metaphor comprehension in reading. *Memory and Cognition*, *2*, 558–567.

Jaarsveld, H., Dijkstra, T., & Hermans, D. (1997). The detection of semantic illusions: Task specific effects for similarity and position of distorted terms. *Psychological research*, *59*, 219–230.

Jackendoff, R. (1977). *X" syntax: A study of phrase structure*. Cambridge, MA: MIT Press.

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

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kamas, E., Reder, L., & Ayers, M. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory and Cognition*, *24*, 687–699.

Katz, A., & Ferretti, T. (2001). Moment-by-moment reading of proverbs in literal and nonliteral context. *Metaphor and Symbol*, *16*, 193–221.

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, *28*, 375–385.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin and Review*, *7*, 257–266.

Kintsch, W. (2001). Predication. *Cognitive Science*, *25*, 173–202.

Laham, D. (1997). Latent Semantic Analysis approaches to categorization. In M. Shafto & M. Johnson (Eds.), *Proceedings of the 19th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1990). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *105*, 221–240.

Landauer, T. K., Foltz, P., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, *25*, 259–284.

MacDonald, M. (1997). Lexical representations and sentence processing: An introduction. *Language and cognitive processes*, *12*, 121–136.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *224*, 522–523.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science*, 226–228.

Meyer, D., & Kieras, D. (1997). A computational theory of executive cognitive processes and multiple task performance. Part 1. Basic mechanisms. *Psychological Review*, *104*, 2–65.

Myers, J., & O'Brien, E. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*, 131–157.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Noordman, L., & Vonk, W. (1998). Memory based processing in understanding causal information. *Discourse Processes*, *26*, 191–212.

Oakhill, J., Garnham, J., & Vonk, W. (1989). The on-line construction of discourse models. *Language and Cognitive Processes*, *4*, 236–386.

Onishi, K., & Murphy, G. (1993). Metaphoric reference: When metaphors are not understood as easily as literal comprehension. *Memory and Cognition*, *21*, 763–772.

Ortony, A., Schallert, D., Reynolds, R., & Antos, S. (1978). Interpreting metaphors and idioms: Some effects on comprehension. *Journal of Verbal Learning and Verbal Behavior*, *17*, 465–477.

Peleg, O., Giora, R., & Fein, O. (2001). Salience and context effects: Two are better than one. *Metaphor and Symbol*, *16*, 173–192.

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: differences in the time course of retrieval. *Cognitive Psychology*, *21*, 139–155.

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

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

Reddy, M. (1993). The conduit metaphor. In A. Ortony (Ed.), *Metaphor and thought*. Cambridge, MA: Cambridge University Press.

Reder, L., & Kusbit, G. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, *30*, 385–406.

Salvucci, D., & Anderson, J. (2001). Integrating analogical mapping and general problem solving: The path-mapping theory. *Cognitive Science*, *25*, 67–110.

Sanford, A., & Garrod, S. (1998). The role of scenario mapping in text comprehension. *Discourse processes*, *26*, 159–190.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwanenflugel, P. J., & White, C. R. (1991). The influence of paragraph information on the processing of upcoming words. *Reading Research Quarterly*, *26*, 160–177.

Shinjo, M., & Myers, J. (1987). The role of context in metaphor comprehension. *Journal of Memory and Language*, *26*, 226–241.

Traxler, M., Bybee, M., & Pickering, M. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *The Quarterly Journal of Experimental Psychology*, *50*, 481–497.

Tyler, L., & Marslen-Wilson, W. (1982). The resolution of discourse anaphors: Some online studies. *Text*, *2*, 263–291.

van Oostendorp, H., & de Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, *74*, 35–46.

van Oostendorp, H., & Kok, I. (1990). Failing to notice errors in sentences. *Language and Cognitive Processes*, *5*, 105–113.