Seminar: *Human Semantic Processing*
Prof. Dr. Manfred Pinkal and Prof. Dr. Matthew Crocker
Winter Semester 2000/2001

Processing Semantic Anomalies: The Influence of Cohesion on Processing Depth

Student: Garance Paris
8th April 2001

Computational Linguistics
Department 4.7 General Linguistics
University of the Saarland

Contents

Introduction 1

1 Early Studies: The 'Moses Illusion' 2

- 1.1 Rejected Explanations . . . 2
  - 1.1.1 Insufficient Knowledge . . . 2
  - 1.1.2 Cooperation . . . 3
  - 1.1.3 Imperfect Encoding . . . 3
- 1.2 Some Influence Factors on the Occurrence of Illusions . . . 4
  - 1.2.1 Focus . . . 4
  - 1.2.2 Context . . . 5
  - 1.2.3 Phonological Similarity . . . 6
  - 1.2.4 Semantic Similarity . . . 6
- 1.3 A First Tentative Explanation: Partial Feature Matching . . . 7

2 Coherence, Cohesion, and Processing Depth 9

- 2.1 Coherence from the Psycholinguistic Perspective . . . 9
- 2.2 'Conceptual Cohesion' . . . 10
- 2.3 The Role of Linguistic Knowledge . . . 10
- 2.4 How Cohesion Influences Processing Depth . . . 11

3 The Influence of Cohesion: Empirical Work by Barton and Sanford 12

- 3.1 Test Material: The Survivors Problem . . . 12
- 3.2 Experiment 1: Influence of Word Meaning Stronger than Presuppositions . . . 13
  - 3.2.1 Design and Method . . . 13
  - 3.2.2 Evaluation and Results . . . 13
- 3.3 Experiment 2: Absence of Influence of Syntactic Complexity . . . 14
- 3.4 Experiment 3: Partial Processing of Noun Phrases . . . 14
- 3.5 Experiment 4: Unclear Influence of Scenario and Reader Expectations . . . 15
- 3.6 Experiment 5: Scenario Influence Proven, but none of Information Ordering . . . 16
- 3.7 Experiment 6: Question-Relevant Information Strengthens the Illusion . . . 17

Conclusion: Initial Cohesion, an Explanation for Semantic Illusions 19

Bibliography 21

Introduction

Just the same as mistakes are part of everyday life, they are also part of everyday speech... However, it is interesting how in most cases, we do not even notice them: we understand what a speaker means even if his utterance actually does not make sense because it contains some error. This robustness can be observed at all levels of human language processing, be it phonetics, syntax, or semantics. Indeed, on the semantic level, it is the source of many common jokes, in which the speaker asks a non-sense question which deliberately contains an incoherent word, such as

When an aircraft crashes, where should the survivors be buried?

In most cases, people will try to provide an answer to answer to the question, although it is, of course, anomalous: everyone knows you don't bury the survivors of an accident, but its victims!

In the literature, the phenomenon of a person not reporting such an anomaly, but instead answering the question as though it made perfect sense, has been termed 'semantic illusion'. Intrigued by the phenomenon, various authors have studied such illusions, examining empirically what factors may increase or decrease the detection rate of semantic anomalies, and this has resulted in interesting new insights on the way people process human language at the semantic level.

This paper will look at one set of such empirical studies, conducted by Barton and Sanford (1990, 1993)[1](chapter 3), and the explanation of illusions offered by Sanford et al. (1995). However, before I go into the work of Barton and Sanford, I will briefly review in chapter 1 the results of some of their precursors, in order to show which explanations had been excluded early from consideration, which factors were known to influence detection, and what was the very first explanation proposed, namely shallow processing and partial feature matching.

Following this, in chapter 2, I will go on to review the notions of coherence and 'conceptual cohesion', and explain how both notions, which are central to Sanford et al. (1995)'s explanation of semantic illusions, relate to one another. Additionally, this chapter will also present the core ideas begind a model of illusions implememted by Budiu (2001).

Finally in the conclusion I will expose Sanford et al. (1995)'s explanation of the phenomenon, which rests on the notion of cohesion.

, considered carefully,

[1]Unfortunately, the article by Barton and Sanford (1990) was not available to me, so that I refer some of the experiments detailed there on the basis of the overview given in Sanford et al. (1995).

1 Early Studies: The 'Moses Illusion'

In the early studies on semantic anomaly detection, participants were confronted with one-sentence questions containing some deliberate distortion, such as the following one which refers to a well-known biblical story:

How many animals of each kind did Moses take on the Ark?

Surprisingly often, the participants in the experiment fell for it and answered "two", not noticing at all that the speaker had substituted the word *Moses* for *Noah*, and this happened even when the participants had been previously warned against "ill-formed" questions.

This "failure to notice discrepancies in questions or assertions" (Reder and Kusbit, 1991) is often termed 'Moses Illusion' because it was first pointed out and investigated by Erickson and Mattson (1981) using the above 'Ark question'. The phenomenon was then studied in more depth by Reder and Kusbit (1991), using a variety of materials, and has proven itself to be quite robust.

Because the work on the Moses Illusion later influenced the research of Barton and Sanford on the detection of semantic anomalies, I will begin this paper by reviewing in this chapter the results of the above cited authors, and also include here some contributions by further related studies.

I will first go over some possible explanations which were excluded early on and briefly give the reasons why these were rejected. Following this, I will survey in section 1.2 some factors of which it soon became apparent that they have an influence on the rate of occurrence of an illusion. This account, however, is restricted to the main conclusions reached by this early research: the empirical aspects have been omitted, and only a selection of arguments in support of each hypothesis are mentioned.[1]

Finally, a preliminary explanation offered by Reder and Kusbit (1991), which made way for future work by Barton and Sanford, will be exposed in section 1.3.

1.1 Rejected Explanations

1.1.1 Insufficient Knowledge

The very first, and obvious, explanation of the phenomenon is that what looks like an illusion to us, in fact may not really be one: in the case of the Ark question, if participants only had a vague knowledge of Bible stories, they might only remember that "some great Bible figure" took animals on an ark, not who exactly it was.

To make sure such "false" illusions did not mix into the data, Erickson and Mattson (1981) tested the knowledge of their participants after the experiment to ensure they could have

[1] This part of the material, which was presented by Ruth Fuchss in the seminar, is the object of her paper on the Moses Illusion (see `http://www.coli.uni-sb.de/~fuchss/Ruth/Uni/Arbeiten/HSP-ausarbeitung.ps.gz`).

noticed the error, and excluded cases in which the participant did not have the necessary knowledge to detect the distortion.

As for Reder and Kusbit (1991), in one of their experiments they required that the participants study the facts in their correct version before hand, and even, in a variant of the experimental design, that the participants learn these facts off by heart. Seeing that this manipulation did not influence the frequency of illusion, Reder and Kusbit concluded that illusions could not be due to insufficient knowledge on the part of the participants.

However Reder and Kusbit went on to make an interesting point. They remarked that Bredart and Modolo (1988) have shown that Belgian participants will easily fall for the illusion when the discovery of America at the end of the 15th century is attributed to Magellan instead of Columbus, and thus argue that, even if insufficient knowledge is not an explanation for the phenomenon, the knowledge state of the participant probably still has some kind of influence, because it would seem highly unlikely that American participants would confuse *Magellan* with *Columbus*.

1.1.2 Cooperation

When their attention is called to the fact that the question was distorted, some victims of the illusion react by alleging they had noticed the error, but understood what the speaker meant and answered consequently, simply ignoring the distortion. This is what would be termed a 'cooperative[2]' reaction, but was ruled out as an explanation of the illusion occurring with most of the participants.

The reason for this was that most participants in the experiments of Erickson and Mattson (1981) (cp. also Barton and Sanford, 1993) were surprised when the distortion was put to them. Moreover, if people were consciously reacting cooperatively, it would be expected that, when expressly requested not to react in this way (an experimental task called 'literal task' by Reder and Kusbit, 1991), they should be able to avoid this type of reaction and instead report the distortion. This was not the case, as was noted by both Erickson and Mattson and Reder and Kusbit.

Reder and Kusbit (1991) also consider the possibility of assuming a weaker version of the Gricean cooperative principle, in which cooperation would not be a conscious process over which people have control. However Reder and Kusbit then bring evidence against this conjecture too, namely the fact that some substitutions are always noticed, e.g. *Nixon* in *How many animals did Nixon take on the ark?*, for indeed under this theory, there would be no explanation for people's not reacting cooperatively in the *Nixon* case, but well in the *Moses* case.

1.1.3 Imperfect Encoding

One last explanation excluded by the early research postulated that victims of the illusion might fail to correctly encode the distorted word in the question sentence, i. e. that people may simply not hear the distorted word, and that readers may be skipping over the critical word during reading.

[2] Language philosopher H.P. Grice's 'cooperative principle' assumes that communication requires cooperation between speaker and hearer. Among other things, his conversational maxims posit that participants in a conversation make relevant contributions and do not say things which they know are not true. Thus, faced with a distorted question, the hearer would on the one hand assume the distortion was a mistake (since the speaker is assumed to be truthful), and on the other hand, he would answer in the way he thinks the speaker expected him to (relevance), instead of vacuously pointing out the mistake.

This was rejected by Erickson and Mattson (1981) firstly on the grounds that the illusion also occurred when the question is read out loud by the participants, and secondly because when asked to repeat the question after answering it, the participants repeated it exactly as it had been presented to them, including the distortion. Thus in both cases the distorted word must at least have been correctly perceived at the phonetic level. Nevertheless it could still be imagined that correct representations of the word might not be passed on to the syntactic or semantic levels of processing.

Evidence against this came from latency data gathered by Reder and Kusbit (1991). If the illusion were due to poor encoding, and assuming that reaction times reflect the number of words processed, one would suppose that, when the illusion is detected in the literal task, the extra processing of the distorted word would increase the latencies in comparison to cases in which the distortion goes unnoticed. However, whenever the participants did detect the distortion in the literal task, it did not take them longer to do so, and consequently this hypothesis too had to be rejected.

1.2 Some Influence Factors on the Occurrence of Illusions

In addition to leading to the rejection of a number of hypotheses, the early research on the Moses Illusion also led to the identification of some factors that seem to influence the frequency with which people are subject to an illusion.

1.2.1 Focus

The first of these factors was that in the questions that were presented to the participants, the distorted word fell into the background of the sentence, i. e. the part that would normally build the presupposition[3], whereas another constituent, namely the topic of the wh-question, was the focus of the sentence, thereby drawing the participant's attention away from the erroneous word: e.g, in the Ark question, the focus is on *how many*, causing the remainder of the sentence, including *Moses*, to be less prominent.

Picturing that this could have an influence on the occurrence of illusions, Erickson and Mattson (1981) decided to test their idea by turning the questions of their material into statements (*Moses took...*) to be rated as true or false by the participants. They posited that this more neutral sentence type would put the focus on the subject, or at the very least take it away from *how many*. When they observed that the illusion also occurred with these statements, they then concluded that focus could not be the source of the illusion.

To this one should add that, as mentioned in section 1.1.2, some substitutions are always noticed, such as in *Who was the first man to walk on the sun?* or when *Noah* is replaced by *Nixon* in the Ark question. This also seems to indicate that focus cannot be a satisfying explanation for illusions, for if simply diverting the participant's attention from the distorted word sufficed to produce illusions, these should then also occur equally with all nouns.

However, the conclusion of Erickson and Mattson (1981) did not exploit the numeric tendencies in their results: even if the illusion continues to take place with statements, the data also showed the illusion to occur much less often with the statements than with the questions,

[3] A presupposition is the information in a sentence that the speaker assumes to be shared by the hearer. In the case of the Ark question, the speaker assumes that Moses' taking some animals on the Ark is a shared belief; the number of the animals however is assumed not to be shared: either the speaker does not know and he wants to learn it from the speaker, or the speaker knows and he thinks the hearer doesn't, and wants to test him.

so that one might actually maintain that focus, while not being an explanation by itself, does have some influence on the frequency of the illusion.

The influence of focus on the occurrence of illusions was then further investigated by Bredart and Modolo (1988), who criticised the statements used by Erickson and Mattson (1981) for being unclear as to the focused element: if the statements had been spoken, the focus would have depended on the intonation, but since the participants read the statements, it is unclear whether, in *Moses took two animals of each kind on the Ark*, the readers interpreted the focus to be on *Moses* or on the number of animals.

Seeking a syntactical device which would make the focus of the sentence unambiguous, Bredart and Modolo conducted a new experiment using it-clefts[4]. They compared two different cases, one in which the distorted word was part of the if-cleft, and one in which some other constituent was cleaved. Their material was composed of French sentence pairs built along the following model:

It was Moses who took two animals of each kind on the Ark.
It was two animals of each kind that Moses took on the Ark.

The results showed that an illusion is much more likely to occur with sentences of the second type, i. e. when the focus of the sentence is on a constituent other than the one containing the distorted word, which again confirmed the influence of focus as a factor of the illusion.

1.2.2 Context

Another factor that was highlighted by the experiments on the Moses Illusion is the influence of context. Erickson and Mattson (1981) were the first to note, in the discussion of their experiment on focus (see previous section), that the illusion also occurs when the inconsistent word is the first in the sentence, thus showing that a preceding context is not absolutely necessary to induce an illusion.

This remark from Erickson and Mattson constituted a first refutation of an implicit hypothesis which was later formulated by Reder and Kusbit (1991), namely that in a question, if there were enough information before the distorted word to actually answer the question, the participants might not process the distorted word as completely, thus leading to an increase in the frequency of the illusion.

However, since Erickson and Mattson's observation did not rely on quantitative measures of any kind, the idea was taken up again by Reder and Kusbit, who investigated whether the illusion would show up more frequently when the distorted word came late in the sentence. Their experiments showed that this was not the case, which led them to reject an influence of word order.

Yet Reder and Kusbit also showed that, if word order does not have an influence on the illusion, the number of content words associated with the answer may have one: in the literal task, reactions times increased in questions where a greater number of content words were associated with the answer, indicating that the task was made more difficult.

This indicates that when we speak of the influence of context, the two aspects of word order on the one hand, and of amount of contextual information on the other hand should be dissociated, for whereas the number of content words may have an influence, word order does not, as long as the number of content words is kept constant.

[4] A cleft construction can be derived from a non-cleft construction by introducing one of its elements, say X, in a clause having the form *It <be> X*, and by turning the rest into a relative clause, as when deriving *It was John [that we saw]* from *We saw John.*

The influence of context and word order were later investigated in more detail by Barton and Sanford (1990, 1993), who arrived at similar conclusions. These further experiments of Barton and Sanford on context and word order will be the object of sections 3.5 and 3.6 of this paper.

1.2.3 Phonological Similarity

In an attempt to explain why the replacement of *Noah* by *Nixon* in the Ark question strikes the reader more than *Moses* (see section 1.1.2), the two last factors that were considered by the research on the Moses Illusion were phonological and semantic similarity of the correct and substituted words.

Indeed, although the experimental findings of Erickson and Mattson (1981) have shown that the participants do encode the distorted word on the phonological level (see section 1.1.3), it remains that some people react to the Ark question by saying they did not hear the speaker correctly and mistook /'moʊzəz/ for /'noʊə/, a fact which could find its explanation in the phonological features shared by the words: both have two syllables, the first of which is stressed, and begin with a nasal, followed by the same diphthong (/oʊ/).

In order to test for an influence of phonological similarity, Erickson and Mattson (1981) compared the frequency of the illusion with different distorted versions of a same sentence, in which the inconsistent words ranged from most phonologically similar to less phonologically similar to the correct word. The variables they manipulated were the first and last sounds of the nouns, their number of syllables, and the position of stress, all the while attempting to keep semantics as constant as possible by restricting possible substitutions to Old Testament patriarchs.

Their results did not support an influence of phonological similarity. However, it must be noted that the scale used to estimate similarity was very coarse, and that the assumption that using all patriarchs keeps the semantics constant is a fairly bold one, so that in the end, phonological similarity cannot really be completely excluded as a factor of influence on the illusion.

1.2.4 Semantic Similarity

But if we have no hard evidence either for or against an influence of phonological similarity, semantic similarity certainly has been shown to act upon the rate of occurrence of the illusion.

On the assumption that the meaning of words is complex and decomposable into sets of semantic features[5] below word level, and that human semantic processing is sensitive to these features (as shown in Smith et al., 1974, cited by Barton and Sanford, 1993), Erickson and Mattson (1981) supposed that the reason why *Moses* is so easily confused with *Noah* in the Ark question may be that both nouns have many common semantic features. Should this be the case, it would also explain why *Columbus* would not seem similar to *Magellan* for Americans, but may seem similar to Belgians (see section 1.1.1), because the differing features between *Magellan* and *Columbus* would be not be as salient for the latter.

As with phonological similarity, they explored this hypothesis by comparing the frequency of the illusion when different inconsistent names were inserted into the same sentence, and established that indeed, the more the semantic features of the incorrect word and the correct one overlap, the more frequent the illusion is. Such an experiment involving "good" and "bad" distortions (distortions in which correct and distorted words are very similar vs distortions in

[5]For example, the noun *Moses* would be made up of features such as human, male, biblical character of the Old Testament, led his people out of Egypt, received the Ten Commandments from God, etc.

which both are less similar) was also later conducted for the literal task by Ayers et al. (1996, cited by Budiu and Anderson, 2000), who obtained similar results.

This finding was later confirmed by further experiments by van Oostendorp and Kok (1990) and van Oostendorp and de Mul (1990). Whereas Erickson and Mattson (1981) relied only on their intuition to determine to what extent the different nouns they used as substitutes were semantically similar to the correct words, van Oostendorp and Kok and van Oostendorp and de Mul used attribute generation tasks[6] in pre-experiments to establish the semantic features of possible substitutions. These sets of features were then used to rank the incorrect words as more or less similar to the correct ones and make predictions as to the frequency of the illusion, predictions which were corroborated by the comparison experiments similar to the one by Erickson and Mattson.

1.3 A First Tentative Explanation: Partial Feature Matching

The early work on semantic anomalies, which was centred on the Moses Illusion, led to the conclusion that obviously an erroneous word in a sentence is not fully processed, but instead only partially analysed. At the time this research was conducted, this conclusion seems to have constituted a fairly novel insight on human semantic processing, for Erickson and Mattson (1981) mention it was often an accepted opinion that sentences are fully checked for consistency – an assumption which appears unwilling to give way easily, since much later, it is mentioned again by Barton and Sanford (1993), only to be refuted once more.

Erickson and Mattson having shown that the source of semantic illusions is shallow processing, Reder and Kusbit (1991) then developed a tentative explanation of the phenomenon, namely that not every feature of the concept in memory and the word in the sentence are matched: instead if the features of the substituted word heavily overlap with those of the correct word, there is a good global fit of the inconsistent word in the sentence, and as a consequence not all features are tested, and the distortion goes unnoticed.

Erickson and Mattson see in shallow processing a matter of necessary efficiency of the human semantic processing mechanism: given that the meaning of most words is made up of large sets of semantic features, an exhaustive check of all features may simply not be realistically feasible.

Furthermore, if we postulate a model of lexical semantics in which some features are inherent to the words and others nonessential (as proposed by Smith et al., 1974, cited by Barton and Sanford, 1993), then it would very often be the case during semantic processing that some of the latter would fail to fit the meaning of the sentence, so that interrupting the checks at some point on the scale between essential and nonessential features would in most cases be beneficial. Put differently, it would not be advantageous for every slight mismatch to produce a global failure, whereas we would instead want the mechanism to fail only in the case that very few or no semantic features fit.

And if we consider what the human semantic processing mechanism has to deal with in everyday spoken language, there are a number of reasons which make such an approach seem very reasonable: first, it is unlikely that a question will exactly match the memory representation, so people would have to be tolerant of small mismatches; second, most errors are unintentional; third, speed is essential to the semantic processing mechanism; and fourth, partial matching usually works, and there are safeguards built into the system allowing large discrepancies to be

[6] A task which consists in the participants' having to generate as many attributes as possible which they judge important or typical of the word involved.

detected readily, so that it would be sensible to let the normal processing mode be as effortless as possible.

2 Coherence, Cohesion, and Processing Depth

In the previous chapter, we saw how Erickson and Mattson (1981) and Reder and Kusbit (1991) established shallow processing as the source of semantic illusions. However, the question of what could possibly drive such a type of processing remained.

A proposal to answer this aspect of the question is made by Barton and Sanford (1993) and Sanford et al. (1995), who consider cohesion establishment to be the mechanism guiding shallow processing. Their argument is based on a definition of cohesion provided by van Oostendorp and den Uyl (1984)[1], which will be exposed in section 2.2.

Following this, Barton and Sanford and Sanford et al.'s idea – briefly put, that if a whole text or sentence if very cohesive, then a small mistake will not be noticed by the hearer – will be the object of section 2.4, before chapter 3 then goes on to present their empirical work.

2.1 Coherence from the Psycholinguistic Perspective

But before we proceed to cohesion and its influence on processing depth, first another notion must be briefly defined, one which is central to human semantic processing in general, and thus to van Oostendorp and den Uyl (1984)'s analysis: that of coherence.

In everyday language, when we say that a text is coherent, we usually refer to a property of the text itself, meaning that it is orderly and thus understandable. However in psycholinguistics, the word coherence also has a another meaning. Assuming that text comprehension requires the construction of a mental representation of the text, coherence is seen as a property of the result of this cognitive process, and it is this mental representation, then, which must be "well-structured" (Sanford and Garrod, 1994) in order for us to understand the text – we say it must be '**coherent**'.

It is in this sense that Sanford et al. (1995) define coherence to be a property of a text in which the meaning of the parts can be related through specific semantic relationships in such a way that the text builds a consistent whole (i. e. one in which there are no logical or semantic contradictions), which can then be assessed as true or false.

Building on this view of text understanding, it is then commonly assumed that, in order to guarantee a consistent mental representation, we conduct thorough checks on the acceptability of a potential filler for a thematic role[2], by matching the filler's meaning to selection restrictions

[1]Unfortunately, this paper was not available to me, so my information is based on the review of the main ideas given at the beginning of van Oostendorp and Kok (1990).

[2]Semantically, all verbs denote some kind of relation between entities. The number of these entities and the role they play in the situation described by a sentence is determined by the semantic definition of each verb, and these 'slots' have to be filled accordingly for the sentence to be well-formed. For example, the verb *hit* requires someone who hits (the agent) and someone who is hit (the patient). The names the filling entities are given – agent, patient, instrument, goal, etc. – reflect their semantic relationship to the verb, but are independent of grammatical encoding (in *John was hit by Bill* the agent is the object), and are called its

on the role slot. Thus for instance, because our semantic memory tells us that the filler of the patient role of the verb *eat* should have a feature *edible*, a sentence such as *John ate his dinner* will be considered coherent, whereas *John ate a house* will be recognised as anomalous.

2.2 'Conceptual Cohesion'

However, if the definition of coherence given in the previous section is a familiar one in linguistics, van Oostendorp and den Uyl (1984)'s definition of cohesion is somewhat different from common linguistic usage, in which it is usually considered to be the surface realisation[3] of the underlying semantic coherence.

According to van Oostendorp and den Uyl, during semantic processing people continuously monitor what they call the 'conceptual cohesion' of the mental representation of a text (abbreviated as 'cohesion' in the following). This depends solely on information which is immediately available, excluding the making of inferences, and is higher when concepts in memory are semantically similar (as shown in section 1.2.4), as well as when they are highly related due to world knowledge.

Indeed, Sanford et al. (1995) suggest that during the first stages of processing, in which the cohesion of a piece of discourse is evaluated, the text is matched to a situational description in memory which serves to provide relevant background information, and if the elements of the utterance are sufficiently similar to those of the scenario in memory, then a high degree of cohesion is obtained.

On the whole, one could say that instead of relating the parts of a text through real, sound semantic relations as in the case of coherence, cohesion requires the filler of a thematic role only to be "in some way" (for instance by its linguistic similarity or due to world-knowledge) relevant to the domain of the verb, in a weak sense. This is then enough to establish a "virtual" semantic relation, which would be detected as flawed if more processing were to take place, but suffices to give the reader an impression of understanding (Sanford et al., 1995).

2.3 The Role of Linguistic Knowledge

The view that semantic processing is based on the matching of utterances to scenarios in our long-term memory is at the core of the model implemented by Budiu (2001) in her doctoral thesis, model which accounts for several empirical phenomena, among which the Moses Illusion and metaphor understanding. Budiu argues that this partial matching of the features of the words in the utterance with those of the concepts which are part of the scenario selected in our background knowledge makes for a robust semantic processing mechanism which allows the understanding of metaphors, but at the same time leads to semantic illusions.

Her model proceeds as follows: given the meaning of the first word, it first searches for a candidate context in its long-term memory, then matches incrementally each of the first few words of the utterance to this candidate, and if up to a certain point there is no reason to reject the candidate, the model then achieves a high "degree of confidence" (comparable to cohesion in Barton and Sanford's terms) that the scenario is correct and subsequently interprets as much

thematic roles. In addition, fillers of a thematic role slot must meet certain conditions which depend on the verb. These conditions, which may for example require a filler to carry a certain semantic feature, are called **selection restrictions** on the slot, because they exclude some potential fillers (see example in main text).

[3] Some means to establish cohesion in a text are for example pronouns, determiners, lexical repetitions, ellipsis and connectives (among others).

as possible further words so that they fit in with the context it has found (Budiu and Anderson, 2000). This results in background knowledge influencing the understanding of new words in the sentence, and may in some cases lead to semantic illusions.

2.4 How Cohesion Influences Processing Depth

But if primary processing is based on similarity between the utterance and background knowledge, excluding the computation of the exact of relations between words in the sentence, when are these then computed? van Oostendorp and den Uyl (1984) oppose the initial phase of processing in which the degree of cohesion is evaluated with a second more demanding phase, during which in-depth comprehension is achieved through the construction of a coherent mental representation of the discourse.

Sanford et al. (1995), too, agree that semantic processing consists of two steps, and Barton and Sanford (1993) mention work by Sanford and Garrod (1989) and Reder (1987), in which different phenomena of human semantic processing, such as anaphoric reference resolution and question answering have been explained with processing mechanisms comprising two levels. One idea is common to all these models: that the second phase, which is more detailed, effortful, and time-consuming, does not take place in cases in which initial conceptual cohesion is high.

Here an analogy is proposed by Barton and Sanford (1993) with the processing aspects of Smith et al. (1974)'s model (mentioned in section 1.3), which, although developed in another context than the Moses Illusion, can be drawn on to explain how the overlap of semantic features, evaluated during the first phase of processing, would determine the extent of subsequent processing: if the overlap is very high, then the concepts would be considered to match, if it is very low, the concepts would be recognised as different, if it is somewhere in-between, a second, more logical and systematic, phase of processing would be required.

Extending this analogy now to a two-level processing mechanism intended to explain semantic illusions, a further parallel emerges: it is only if cohesion is neither very high, leading to acceptance of the sentence as well-formed, nor very low, leading to a fast rejection of the sentence as anomalous, in other words only if a decision cannot be made after the first phase of processing, that processing may carry on into a second phase, in which an actual verification of coherence would take place in a detailed and time-consuming way.

Indeed a processing mechanism in which a first stage that is intuitive and fast but prone to error would determine whether to go into a more effortful, but for that matter error free, second stage, would explain why semantic anomalies are allowed to slip by in cases in which initial cohesion is high, although upon closer inspection, people would definitely notice the inconsistency.

However, this leads us to reject the assumption of thorough testing of the fit of a filler to a slot, which was mentioned in section 2.1: instead, the depth of the checks on coherence seem to depend on the degree of cohesion established during the first phase of processing.

In the end, the general picture is one in which it seems as if the human comprehension system were assuming coherence as a default, conducting in most cases only a cursory analysis, and as if there were some sort of trade-off between cohesion and coherence, leading to mental representations which are perhaps coherent, but also to a certain extent incomplete (Sanford et al., 1995).

3 The Influence of Cohesion: Empirical Work by Barton and Sanford

Admitting that it is the partial matching and shallow processing which lead to semantic illusions, and that these are driven by cohesion as it is defined by van Oostendorp and den Uyl (1984), it is natural to ask oneself next what factors control the degree of cohesion and consequently the depth of semantic analysis.

It was in order to investigate this that Barton and Sanford (1990, 1993) conducted the series of experiments described in the following chapter, which all build on the assumption that, if high cohesion leads to non-detection of semantic anomalies, then, more generally, by studying the detection rate of anomalies under diverse conditions we can somehow obtain an indirect measure of the degree of cohesion in each case (Sanford et al., 1995).

3.1 Test Material: The Survivors Problem

For their experiments, Barton and Sanford chose another paradigm than the one used in previous studies of the Moses illusion. Indeed, these had used as test material series either of one-sentence questions which the participants were expected to answer, or of assertions which they where supposed to rate as true or false. However, this was not a very "natural" experimental setting, and it cannot be excluded that the participants' having to answer or judge a long list of questions/assertions might have influenced their reactions.

A more natural setting was the one chosen by Barton and Sanford, who used a single test item, actually a common joke, in which the hearers are told a story and then asked their opinion, but the question actually does not make sense because a distorted word was introduced in the narrative:

> *There was a tourist flight travelling from Vienna to Barcelona. On the last leg of the journey, it developed engine trouble. Over the Pyrenees, the pilot started to lose control. The plane eventually crashed right on the border. Wreckage was equally strewn in France and Spain. The authorities were trying to decide where to bury the survivors.*
>
> *What should the authorities do?*

Of course, the point here is that survivors most certainly do not need to be buried, but very often participants did not notice this, a fact which Barton and Sanford attributed to incomplete processing of the word *survivors*, following in this Reder and Kusbit (1991)'s explanation of the Moses Illusion.

3.2 Experiment 1: Influence of Word Meaning Stronger than Presuppositions

Using the aircrash scenario, Barton and Sanford conducted a first series of experiments by replacing the word *survivors* by several variants. In one of these, they investigated whether the anomaly detection rate would be higher when the erroneous noun actually carried the inconsistent semantic feature *is alive* than when this feature was only presupposed.

Indeed, we saw in section 2.2 that, according to Sanford et al. (1995), the degree of cohesion depends on how well a word fits into the context provided by background knowledge, and that if the fit is good, then not all details of the semantic representation of the sentence may be computed, leading to an only partial verification that the word's meaning really complies with the role slot it fills.

In this view, it is then natural to assume that among all features making up the meaning of a word, the most prominent would be checked, so that given several erroneous words which all "fit in" equally well, we would expect those which actually carry an inconsistent feature as part of their meaning to be more readily spotted than those which only presuppose an inconsistent feature, since the inconsistent feature would be more prominent in the first case than in the second (Barton and Sanford, 1993).

3.2.1 Design and Method

Three nouns, *(the) injured*, *(the) wounded* and *(the) maimed* were thus selected for substitution in the aircrash passage, and a pre-experiment established the validity of the assumption that these only presuppose a feature *is alive*, but that *survivors* actually carries the same feature as part of its meaning. Using these three nouns, three experimental conditions were constructed:

1. the unmodified aircrash passage;
2. the "injured-group", which was composed of the aircrash scenario in which the noun *survivors* was replaced by the three nouns *injured*, *wounded* and *maimed*;
3. and a group in which the same three nouns were prefixed with the adjective *surviving*, so that the noun phrase carried the feature *is alive* as part of its meaning.

It was expected that the detection rate for the first and third conditions, in which *is alive* is part of the meaning of the noun phrase, would be higher than in the second condition, in which *is alive* is only presupposed. The participants were shown first another text without anomaly as an exercise, then the aircrash passage in one of the three conditions, and were asked to write down their answer to the final question *„What should the authorities do?"*. They were then debriefed, and the results of the debriefing were taken into account when classifying them into detectors and non-detectors.

3.2.2 Evaluation and Results

In those cases in which the written answer explicitly mentioned the anomaly, the participants were classified as "detectors". Whenever this was not the case, the participants were subsequently asked whether they had noticed anything odd about the text. In a small number of cases, the answer was yes, and the answer was considered to be a cooperative response and the participant classified as a detector. However in most cases, the inconsistency had to be

Figure 1: Detection rate in experiment 1

pointed out to the participants, which reacted with surprise and admitted not having noticed anything – these participants were then classified as "non-detectors". Finally, the participants were asked if they had encountered the aircrash story before in any form, and any participants that had were excluded from the analysis.

Overall the anomaly was detected in only 30 % of the cases. Figure 1 shows that there was no statistical difference between the three items in neither the second nor the third condition, but that the difference between the first and second conditions was relevant, thus confirming the predictions, and moreover, that the introduction of the adjective in the third condition had the expected effect of increasing the detection rate to the level observed in the first condition.

3.3 Experiment 2: Absence of Influence of Syntactic Complexity

In another experiment, Barton and Sanford (1990)[1] replaced the word *survivors* from the original aircrash passage with variants having the same meaning but of differing syntactic complexity. Indeed, it has been shown in earlier studies (Anderson, 1982; Sanford and Garrod, 1981, both cited in Sanford et al., 1995) that more complex noun phrases are more likely to be referred to in continuations than simple ones, suggesting that they are more salient, possibly because they require extra processing, something which should lead to a higher detection rate of anomalies.

From the aircrash passage, Barton and Sanford thus created three experimental conditions by replacing *survivors* by *passengers who survived* and *surviving passengers*. However, the detection rate being 59 % for the simple noun phrase, 43 % for the noun phrase containing a relative clause and 47 % for the one with an adjective, the hypothesis that syntactical complexity may have increase the detection rate of the anomaly had to be rejected.

3.4 Experiment 3: Partial Processing of Noun Phrases

Surprisingly, the results of the previous experiment were actually the opposite of what had been expected: the detection rate in more complex noun phrases in which the modifier (adjective or relative clause) was anomalous was even lower than in the case of the simple noun phrase. This seemed to suggest that modifiers receive very little attention during processing.

In order to investigate this more closely, Barton and Sanford (1993) conducted a follow-up experiment in which they tested to what extent a noun phrase anomalous in its own right, *the surviving dead*, would be detected. The results of the experiment confirmed the expectation that the anomaly would be spotted very rarely: the detection rate of 23 % was much lower than that in the previous experiments where *survivors* had been used.

On an account of semantic processing in which local meaning of the noun phrase would be computed before being integrated into the meaning of the sentence, this would be fairly hard to explain. Instead the observations seem to point to a processing mechanism in which strong top-down constraints due to our world knowledge determine the amount of processing each word is afforded, based on the availability of a part of its meaning early on during processing (Sanford, 1999, p. 329).

[1]This experiment by Barton and Sanford (1990) is reviewed in Sanford et al. (1995).

Figure 2: Detection rate in experiment 4

	Aircrash	Bicycle accident
Survivors	59 %	50 %
Surviving dead	20 %	65 %

In this view, we would thus assume that the reason the anomaly in the aircrash passage is not detected is that the word *survivors* is relevant in the context of an aircrash, so that the details of its relation to *bury* are not computed, and that similarly, it is because noun and adjective in *the surviving dead* both fit well into the context that, just as the soundness of the relation between a verb and its patient is not fully checked, the soundness of the relation between the noun and its modifier, too, is not be completely verified.

3.5 Experiment 4: Unclear Influence of Scenario and Reader Expectations

The results of the experiments exposed in the previous sections confirmed the supposed role of context and background knowledge in semantic processing, which appear to act as strong top-down constraints inhibiting checks on local coherence when the meanings of individual words fit well into the context at hand. If this is true, then it should be possible to show that in a context in which *survivors* does not fit as well because the setting does not imply as strongly the presence of dead victims, more checks on coherence are made, and thus the anomaly is detected more readily.

This is what Barton and Sanford (1990)[1] set out to explore in a following experiment. A text very similar to the aircrash passage was generated, but in which the event was a bicycle accident:

> *There was a cycle race being staged from Milan to Geneva. The leaders were approaching the border of Italy and Switzerland. Suddenly, the leading cyclist got a puncture and lost control of his bike. He fell awkwardly in front of twenty other competitors. The whole pack ended up in a heap right on the border. The wreckage was equally strewn in Italy and Switzerland. The authorities were deciding where to bury the survivors. They couldn't make up their minds.*

In addition to varying the scenario, two variants of the noun phrase were used, *survivors* and *surviving dead*, thus resulting in four experimental conditions. It was anticipated that the detection rate would be higher in the case of the bicycle crash than in the case of the airplane crash, because dead victims, although not being incompatible with a bicycle accident, are not expected by the reader in this context.

The results of the experiment appear in figure 2: the predictions were confirmed in the case of the anomalous noun phrase *surviving dead*, but not in the case of *survivors*, so that this study only provided some evidence for the influence of scenario on anomaly detection, but as strong an evidence as predicted.

Figure 3: Detection rate in experiment 5

	Aircrash		Bicycle accident
	visual presentation	auditive presentation	
Early scenario, late verb	20 % (26 %)	33 %	80 %
Early scenario, early verb	27 % (44 %)	46 %	73 %
Late scenario, late verb	33 % (31 %)	26 %	73 %
Late scenario, early verb	53 % (46 %)	41 %	93 %
Scenario Average	33 % (37 %)		80 %

3.6 Experiment 5: Scenario Influence Proven, but none of Information Ordering

Having established in the previous experiments that an illusion could indeed also be induced with a text, but wishing to gather more data concerning the impact of context on semantic illusions, Barton and Sanford used another design for the following experiments: the critical item was shortened to a one-line question and presented to the participants as the fifth element of a 10-question questionnaire about diverse social issues (e. g. *At what age should it be legal to smoke tobacco?*). The participants were asked to write what they thought about each issue, and after they had filled in the questionnaire, they were debriefed as in the previous experiments.

Using this new design, another experiment on the influence of context on illusions was carried out by Barton and Sanford (1993). First, the assumption that the perceived likelihood of dead victims is high for an aircrash but relatively low for a bicycle accident was verified in a pretest with 16 participants, who were asked how likely they thought dead victims would be in different types of accidents.

This premise having been established, the aircrash and bicycle scenario were again opposed. Additionally, since the hypothesis was that context is responsible for illusions, the order in which the information was presented was also taken into account: the scenario was presented either before or after the clause containing the critical word, and the verb phrase was either passive or active, so as to test for effects of presenting the word *survivors* before or after *bury*. All in all, this resulted in four question types and eight conditions:

1. Early scenario, late verb (passive VP)
   When (an aircraft crashes/a bicycle accident occurs), where should the survivors be buried?
2. Early scenario, early verb (active VP)
   When (an aircraft crashes/a bicycle accident occurs), where should you bury the survivors?
3. Late scenario, late verb (passive VP)
   Where should the survivors be buried after (an aircrash/a bicycle accident)?
4. Late scenario, early verb (active VP)
   Where should you bury the survivors of (an aircrash/a bicycle accident)?

Concerning the ordering of information, the prediction was that, if the semantic processing mechanism is assumed to proceed strictly from left to right, the more information is available before the critical word, the smaller the detection rate should be, so that type 3 sentences should have the highest detection rate and type 2 sentences the lowest, the respective detection rates in type 1 and type 4 sentences depending on the relative impacts of verb placement and scenario.

The second and last columns in figure 3 show the results of the experiment. Overall, the detection rate in the bicycle accident context was much higher, thus confirming that the scenario has an influence on the detection of the illusion. However, no influence of information ordering could be proven because the differences between conditions were not statistically significant.

For this reason, the ordering hypothesis was investigated a second time with further participants, using only the aircrash scenario. The results, which appear between parentheses in figure 3, show a marginal difference between early and late verb placement, but in the opposite direction than the one predicted: the detection rate was slightly higher when the verb occurred before the anomalous word (in active sentences). Although this is not a statistically significant result, should the numerical trend in the data nonetheless be interpreted, it would seem to hint that the critical word is not processed immediately: instead it appears to be processed later, when the whole sentence is available, for apparently, information occurring after the anomalous word may still contribute to illusion effects.

To exclude the possibility of word order making no impact because participants read the sentences more than once, the experiment was then repeated once more with an auditory presentation of the sentences (Sanford et al., 1995). However, the results (see figure) were similar to those obtained in the experiment with the larger sample. Consequently, information ordering was rejected as influence factor, thus confirming the conclusions drawn by Reder and Kusbit (1991), exposed in section 1.2.2.

3.7 Experiment 6: Question-Relevant Information Strengthens the Illusion

As was mentioned in section 1.2.2, it seems that when we speak of the influence of context, the two aspects of information ordering on the one hand, and of amount of available contextual information on the other hand should be dissociated. For this reason, having confirmed that information ordering does not have an influence on the occurrence of illusions, Barton and Sanford then turned to the second of these aspects, and explored whether adding information relevant to the question might influence the frequency of the illusion.

As in the experiment exposed in the previous section, questionnaires were used, but this time the critical item was composed of two sentences: the first a description of the context which contained the word *survivors* at its end, enabling the introduction of a relative clause with supplementary information after the critical word, and the second the question concerning the burial place.

Figure 4: Detection rate in experiment 6

The answers in the previous experiments determined what kind of information is considered to be relevant to where the victims should be buried: origin of the victim (country or town of domicile) or relatives' desire. Four conditions were used:

1. Basic version:
   Suppose there was an aircrash with many survivors.
   Where should they be buried?
2. Question-irrelevant information added:
   Suppose there was an aircrash with many survivors, which happened last week.
   Where should they be buried?
   Variant:
   Suppose there was an aircrash with many survivors, most of whom were gravediggers.
   Where should they be buried?
3. Question-relevant information added:
   Suppose there was an aircrash with many survivors who where mainly from Europe.
   Where should they be buried? (+ variants)
4. Control condition:
   Suppose there was an aircrash with many survivors who were mostly unhurt.
   Where should they be buried?

In conditions 1 and 2, similar detection rates were expected since the added information should not help in anyway to answer the question, whereas in condition 3, a lower rate was predicted and in condition 4 a very high one. The outcome of the experiment is shown in figure 4. The results were those expected: there was no difference between the first two conditions, in which the anomaly was detected on average in 76 % of the cases. In the condition in which information relevant to the question was added, the detection rate was much lower (49 %), and in the control condition, no illusion at all occurred.

These results support the idea of a coherence/cohesion trade-off as mentioned in section 2.4: if early on in processing there is enough information to achieve a high level of cohesion, allowing in this case to answer the question at the end, then further checks on cohesion may not take place, thus leading to an illusion.

Conclusion: Initial Cohesion, an Explanation for Semantic Illusions

In chapter 1 of this paper, we reviewed several factors whose impact on the frequency of occurrence of semantic illusions had been observed in early studies of the Moses Illusion: sentence focus, the amount of contextual information, (possibly) phonological similarity and semantic similarity of the correct and substituted words.

In particular, we saw how the demonstrated importance of this last factor led Reder and Kusbit (1991) to conclude that semantic illusions are due to shallow semantic processing, because only part of the features of the word at hand would be matched with those of the concept in memory during semantic processing.

However, knowing that semantic processing is not always shallow, since some distortions, such as the word *sun* in *Who was the first man to walk on the sun?*, are always noticed, we then asked ourselves what it is that controls the depth of processing: when is semantic processing shallow and when is it not?

Chapter 2 then examined Sanford et al. (1995)'s suggestion that, if human semantic processing consists of two phases, as is suggested by various authors (van Oostendorp and den Uyl, 1984; Reder, 1987; Sanford and Garrod, 1989), then it may be that its overall depth is determined by a quantitative criterion computed in the first phase, namely the degree of 'conceptual cohesion'. A further supposition of Sanford et al. (1995) was then exposed, namely that the evaluation of this initial quantity relies on a mechanism which compares the content words in the utterance with the concepts in a scenario retrieved from our long-term memory, so that the degree of cohesion obtained would depend on how well both series of semantic features match.

Following this, a model implemented by Budiu (2001) in her doctoral thesis on the role of background knowledge was briefly sketched, model that attempts to explain both semantic illusions and the human faculty to understand metaphors with a combination of the above mentioned mechanisms, namely partial scenario-matching, positing that this would yield a robust processing mechanism which would both be tolerant of errors, thus explaining semantic illusions, and allow also the drawing of parallels between comparable elements.

A number of experiments conducted by Barton and Sanford were then covered in chapter 3, which revealed a number of other cohesion-supporting factors. Especially experiments 1 and 5 showed that background knowledge actually has a much broader influence than it was thought until then: whereas under the assumption that cohesion is solely due to scenario-matching, as modelled by Budiu, background knowledge would be restricted to its linguistic aspects (i. e. semantic features), the experiments by Barton and Sanford proved that world knowledge also plays an important role in cohesion establishment, and thus in the occurrence of semantic illusions.

On the whole, it seems as if human semantic processing were very much expectation-driven, i. e. strongly influenced by top-down constraints, so that a variety of early, global processes end up dominating the more local processes of coherence establishment. However, among these many cohesion-building processes, it remains to be determined empirically which operate when and to what extent some may be incomplete.

Thus Sanford et al. (1995) propose a picture of the human semantic processing mechanism in which several processes compete against one another, but in which partial scenario matching would play an important role, so that overall, the extent of the contribution of a given word would depend on many factors instead of just one, thus leading in some cases to the ignorance of semantic anomalies.

Bibliography

Works Cited

Barton, S. B. and Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory and Cognition*, 21(4):477–487.

Bredart, S. and 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*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. Available as technical report no. CMU-CS-O1-148, also from `http://www.cs.cmu.edu/~ralucav/papers/thesis.ps.gz`.

Budiu, R. and Anderson, J. R. (2000). Integration of background knowledge in sentence processing: A unified theory of metaphor understanding, semantic illusions and text memory. In Taatgen, N. and Aasman, J., editors, *Proceedings of the 3rd International Conference on Cognitive Modeling*, pages 50–57, Groningen, The Netherlands. Universal Press. Available from `http://www.cs.cmu.edu/~ralucav/papers/iccm00.ps.gz`.

Erickson, T. D. and Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Meaning and Verbal Behavior*, 20:540–551.

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

Sanford, A. J. (1999). Word meaning and discourse processing: A tutorial review. In Garrod and Pickering, editors, *Language Processing*, chapter 10, pages 301–331. Psychology Press, Hove, UK.

Sanford, A. J., Barton, S. B., Moxey, L. M., and Paterson, K. (1995). Cohesion processes, coherence, and anomaly detection. In Rickheit, G. and Habel, C., editors, *Focus and Coherence in Discourse Processing, Part II*, pages 203–213. de Gruyter, Berlin.

Sanford, A. J. and Garrod, S. C. (1994). Selective processing in text understanding. In Gernsbacher, M. A., editor, *Handbook of psycholinguistics*, chapter 21, pages 699–719. Academic Press, New York.

Smith, E. E., Shoben, E. J., and Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81(3):214–241.

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

van Oostendorp, H. and Kok, I. (1990). Failing to notice errors in sentences. *Language and Cognitive Processes*, 5(2):105–113.

Other References

Anderson, A. (1982). Text comprehension – The influence of temporal information on processing and reading rate. Unpublished PhD thesis, Department of Psychology, University of Glasgow.

Ayers, M., Reder, L. M., and Anderson, J. R. (1996). Accepting false information now and believing it later: Partial matching and false information in the Moses Illusion. Unpublished manuscript.

Barton, S. B. and Sanford, A. J. (1990). Failures to notice semantic anomalies in discourse: Incompleteness of processing in the machinery of cohesion establishment. Paper presented at the Meeting of the European Society for Cognitive Psychology, Como, Italy.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, 19:90–138.

Sanford, A. J. and Garrod, S. C. (1981). *Understanding written Language*. John Wiley & Sons, Chichester.

Sanford, A. J. and Garrod, S. C. (1989). What, when and how?: Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes*, 4:135–162.

van Oostendorp, H. and den Uyl, M. J. (1984). Semantic relatedness effects in text processing. Paper presented at the joint meeting of the British Experimental Psychology Society and the Netherlands Psychonomic Foundation, University of Amsterdam.