DUAL-SYSTEM PROCESSING IN LANGUAGE COMPREHENSION

Jon M. Slack
Open University, England

1 Imagery and Language Understanding

Natural language parsers map linguistic input strings into symbolic, relational structures using syntactic knowledge, semantic knowledge, or both. However, no parser maps such inputs into image codes which implies that people building language understanding systems do not regard image generation and processing as a necessary part of language understanding. Within psychology, on the other hand, some researchers have advocated a strong relationship between image processing and language understanding for the past decade [1]. The status of this view is by no means unequivocal amongst psychologists, and the debate about imagery has littered the pages of many an academic journal. To facilitate your reading of this paper it is worth stating that the work reported here is driven by the belief that image processing is an essential component of language understanding.

2 Using Image Codes in Language Understanding

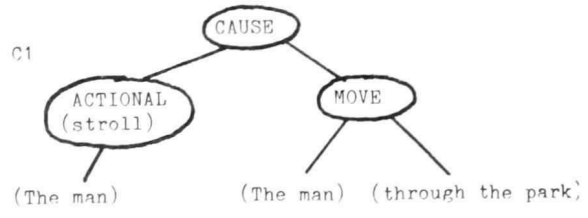
If you ask people to outline the difference in meaning between the words strolling and sauntering they find the task relatively difficult. If you ask them how they arrived at their description, the majority of people state that they formed images of a person strolling, and sauntering, then compared these images generating the best verbal description of the differences between the two they could. This sort of protocol data suggests that it is necessary to process image codes in order to distinguish between the meanings of the two words. That is, at least part of the meaning of the words is represented in some form of image code. If this is the case, then for a natural language understanding system to distinguish between sentences S1 and S2 it needs to make recourse to the image-coded

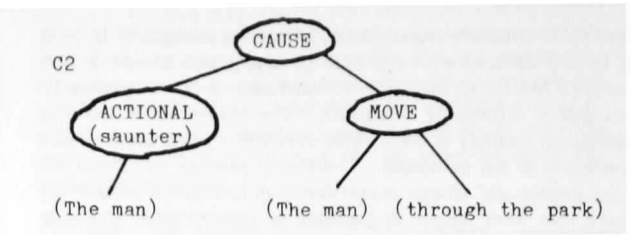
S1 The man strolled through the park.

S2 The man sauntered through the park.

components of the two verbs. This is not saying that an understanding system needs to map the sentences directly into image codes in isolation of other semantic structures. Rather, image processing is an essential component of the total processing involved in building cognitive structures which represent the differentiated meanings of the two sentences.

Most existing language understanding systems parse linguistic inputs into some form of propositional network which represents the conceptual relations corresponding to the meaning of the input [2],[3],[4]. These systems never use non-propositional codes, that is, language understanding is totally contained within a propositional system. Using the work of Norman and Rumelhart [5] as an example system, sentences S1 and S2 would be parsed into the conceptual structures C1 and C2, respectively.



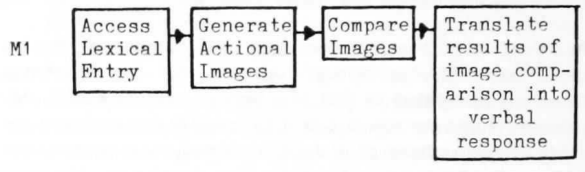


Within this system the two sentences are differentiated in terms of the actional components which represent the physical actions which implement the movement component implicit in the meanings of the verbs. These actional components are associated with the image codes used in distinguishing the verbs.

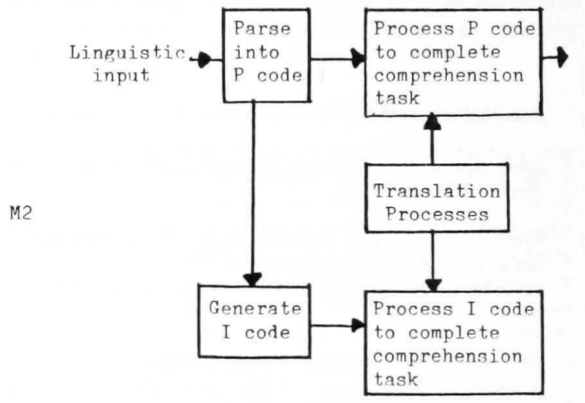
Kosslyn and his fellow researchers [6] have developed a detailed theory of image generation and processing which is backed-up with a working computer simulation. He proposes that numerous cognitive tasks, in partiuclar size-comparison and sentence-verification tasks, involve the processing of both propositional and image codes. The two types of codes and associated processes constitute independent, but connected, processing systems which run in parallel. Kosslyn's theory provides a good foundation on which to build a more adequate model of language understanding incorporating image processing in addition to existing ideas of propositional processing.

3 Model of Language Understanding

The protocol data described in section 2 suggests that the processing underlying the meaning-differentiation task can be accounted for by model M1.



In line with the ideas embodied within Kosslyn's model, a dual-system model of language understanding would have a structure similar to model M2.



In M2 the Propositional code system (P-system) and Image code system (I-system) are not strictly independent in that linguistic inputs are not mapped directly into images. Rather, the I-code is generated as a product of the parsing mechanism which maps the input into a P-code. However, once the two codes are created they are processed separately, but knowledge can be passed from one system to the other by means of the translation processes. Model M1 is accommodated within model M2 by the I-system and translation processes.

But what arguments are there for advocating the addition of an image processing system to existing language understanding models?

Argument 1: the I-system has either sole access to, or more direct access to, knowledge which is

necessary for efficient language comprehension. This knowledge maybe represented only as I-codes. On the other hand, it could be coded within both the I-system and P-system, but more accessible as images.
Argument 2: the I-system is more efficient at processing large amounts of knowledge within specific task domains. For example, I-codes provide a more natural medium for processing knowledge of spatial relations [7]. Further, Waltz [8] has argued that event simulation is an important aspect of language understanding, allowing the comprehension system to make inferences about scenes and judge the plausibility of inputs during parsing. The I-system is necessary for efficient processing of such event simulations.

4 Support for dual processing in comprehension

The imagery value of sentences has been shown to exert a profound influence on comprehension and memory tests [9],[1]. Thorndyke [10] has shown that subjects' imagery ratings for sentences vary inversely with comprehension RTs. Kosslyn's research project team have shown that imagery is used in sentence-verification tasks and for accessing knowledge of physical dimensions of objects. Further experimental evidence in support of Argument 1 is described below.

Experiment 1 Subjects were shown simple subject-verb-object sentences and asked either to read them for understanding (condition 1) or to generate an image corresponding to the meaning of each sentence (condition 2). In condition 1, the comprehension RTs were measured. In condition 2, the image-generation time were measured. Following the presentation of the sentences, subjects in both conditions were given a recognition test. For each target sentence in the test there were two distractor sentences which only differed from the target sentence in terms of the verb. However, the verbs in the distractor sentences were close in meaning to the target sentence verb, as shown below.

Target sentence: The man strolled through the park

Distractors: The man sauntered through the park
The man walked through the park

The results showed that the comprehension times for stimulus sentences are lower than the image-generation times. However, the probability of recognising a target sentence is much higher in the image-generation condition than the comprehension condition. This implies that Ss access information in the image generated for each sentence which allows them to distinguish between a target sentence and the semantically close distractors. Of course, the results are open to more than one interpretation because the difference in recognition probablities between the two conditions could be due to the difference between comprehension and image-generation times. To take account of this possibility, the experiment was repeated in a modified form.

Experiment 2 Subjects in condition 1, rather than just read the stimulus sentences were asked to categorize them according to whether the actions they described involved an object as an instrument of the action. For example, sentence S1 would be classed as NO INSTRUMENT, whereas S3 would be classed as INVOLVING AN INSTRUMENT. The categorization RTs were measured. The rest of the

S3 The chef chopped the vegetables.

experiment was the same as Expt. 1. The results showed that the categorization times were slightly longer than the image generation times, but the recognition probabilities were higher for the image generation condition than the categorization condition. Thus, the results seem to support the conclusions derived from the first experiment.

These experiments seem to imply that Ss access knowledge via images which allows them to differentiate semantically close verbs. However, it would seem that Ss do not access this knowledge in the normal process of comprehension as evidenced by the poor recognition performance in the comprehension condition. This can be construed as evidence against the involvment of imagery in language understanding, but if the I system takes longer than the P system to process an input then the I system may be used as a back-up processing capacity which is only resorted to when P system processing has failed, or proved inadequate. Evidence for this notion comes from work on the comprehension of metaphor which shows that people use imagery to work out the meanings of figurative inputs [Note 1]. The experiment below provides further support for this idea.

Experiment 3 - Subjects had to judge whether stimulus sentences were semantically acceptable or not and the decision RTs were measured. The sentences were constructed so that they would be easy to classify, in either direction, or difficult to classify leading to a bi-modal distribution in decision times. The same sentences were used in an image generation task and the response times were measured. As expected, the results showed a strong bi-model distribution for the decision RTs; Ss tended to be either quick or slow at making judgements. The important findings, however, were that (i) there was no significant difference between the image generation times for fast-RT sentences and slow-RT sentences, and (ii) the correlation between image generation time and decision RT was high for slow judgements, but low for fast judgements. These results imply that when Ss have difficulty in judging the semantic acceptability of a sentence they base the judgement on image coded data rather than knowledge stored in propositional form which is the case for more straightforward judgements. The evidence presented in this section gives weight to the theory that image processing is an important component of language understanding. Specifically, the data show that (i) image codes represent knowledge not directly available to the P-system, but which needs to be accessed in certain comprehension tasks, and (ii) image processing is often employed to solve linguistic problems.

5 Building a complete Language Understander

To build a flexible language understander with powerful problem solving capabilities it is necessary to augment existing P-system parsers by adding image generation and processing routines. The I-system and P-system would function in parallel, with the latter producing a skeletal parse adequate for some comprehension tasks. The I-system would generate a richer (semantically) parse, but over a longer time course. When the P-system output satisfies the requirements of the comprehension task image processing is terminated, uncompleted, and the understander passes on to the next input. If, on the other hand, the output from the P-system is inadequate for the comprehension task, then the richer I-system output is used to solve the problem. At any processing stage after generation, image codes can be inspected and potentially translatable elements of a code can be passed over to the P-system in order to generate a verbal response. The I-system has no direct output mechanism.
Kosslyn's computer simulation of his I-system theory [11] provides a good basis for a working image processing component which can be knitted

into a P-system parser, although it needs to be extended to take account of dynamic images and event simulations. It is not clear how his system would generate an image of a man strolling, as he does not specify how it is possible for components of an image to move relative to each other in a meaningful way. To implement an event simulation of sentence S1 it is necessary to access and run over time an ordered sequence of "key" image frames which represent the actional component of the verb 'stroll'. If you imagine a continuous motion sequence to be broken down into a series of static images, then the key images are those which have a high information content in that they distinguish between the meanings of different verbs. Such images might correspond to a discontinuity in the movement sequence, or a change of direction of motion of an image element. These sets of key image frames should be abstract enough to take different objects/agents of an action as the content of the image. The actional component of the verb within the P-system would be linked to its corresponding key image frame sequence within the I-system, and when the actional component is accessed during parsing the key image frame sequence would be run in the I-system to generate the event simualtion. At the same time, the other P-system elements accessed during parsing, corresponding to the arguments of the verb, would activate object images to form the content of the event simulation. These object images would be interpreted by the key image frame sequence to produce the dynamic image. These ideas have only been discussed in the context of representing the verbs of movement, but they could be generalised to provide a basis for generating the full range of dynamic images and event simulations.

6 References

[1] Paivio, A. Imagery and Verbal Processes. Holt, Rinehart and Winston, New York, 1971.

[2] Schank, R.C. Identification of conceptualizations underlying natural language. In R. Schank and K. Colby (eds.), Computer models of thought and language. Freeman, San Francisco, 1973.

[3] Wilks, Y.A. Grammar, Meaning, and the Machine Analysis of Language. Routledge, London, 1972.

[4] Rumelhart, D.E. and Levin, J.A. A language comprehension system. In D.A. Norman and D.E. Rumelhart (eds.), Explorations in cognition. Freeman, San Francisco, 1975.

[5] Norman, D.A. and Rumelhart D.E. Explorations in cognition. Freeman, San Francisco, 1975.

[6] Kosslyn, S.M. Image and Mind. Harvard University Press, Cambridge, Mass., 1980.

[7] Boggess, L.C. Computational interpretation of English spatial prepositions. Unpublished Ph.D. dissertation, Computer science department, University of Illinois, Urbana, 1978.

[8] Waltz, D.L. Generating and understanding scenen descriptions. In Joshi, Sag and Webber (eds.), Elements of discourse understanding. Cambridge University Press, London, 1980.

[9] Paivio, A, and Begg, I. Imagery and comprehension latencies as a function of sentence concreteness and structure. Research Bulletin No. 154, department of Psychology, University of Western Ontario, 1970.

[10] Thorndyke, P.W. Conceptual complexity and Imagery in Comprehension and Memory. Journal of Verbal Learning and Verbal Behavior, 14, 359-369, 1975.

[11] Kosslyn, S.M., and Shwartz, S.P. A simulation of visual imagery. Cognitive Science, 1, 265-295, 1977.

Reference notes

Note 1. Slack, J.M. Metaphor Comprehension A special mode of language processing? Paper presented at the 18th. annual meeting of the Association for Computational Linguistics, Philadelphia, June 1980.