Making Mathematical Connections Through Natural Language: A Computer Model of Text Comprehension in Arithmetic Word Problem Understanding

Mark D. LeBlanc
Department of Computer Science
University of New Hampshire
Kingsbury Hall, 33 College Road
Durham, NH 03824-3591
mdl@cs.unh.edu

Sylvia Weber-Russell
Department of Computer Science
University of New Hampshire
Kingsbury Hall, 33 College Road
Durham, NH 03824-3591
swr@cs.unh.edu

Abstract

Understanding arithmetic word problems involves a complex interaction of text comprehension and mathematical processes. This work presents a computer model of the hypothesized processes that are required of a young student solving arithmetic word problems, including the processes of sentence-level reading and text integration. Unlike previous computer simulations of word problem solving, which neglect the early stages of text processing, this model forces a detailed consideration of the linguistic process, which is being increasingly recognized as a primary source of difficulty. Experiments were conducted to isolate critical text comprehension processes. Children's probability of solution was analyzed in regression analyses as a function of the model's text comprehension processes. A variable measuring the combined effects of the load on working memory and text integration inferences accounted for a significant amount of variance across four grade levels (K-3). The results suggest new process-oriented measures of determining why a particular word problem may be difficult, especially for young students. An implication for education is the potential for a difficulty-differentiated network of problems that includes a multiple number of rewordings for each "traditional" problem wording as an aid for classroom assessment and future computer-based learning environments.

Introduction

Researchers in cognitive science and mathematics education have recently emphasized that it is not only mathematical problem solving abilities that are at the root of children's difficulties with mathematical word problems, but also *linguistic* difficulties. That is, many difficulties occur at the stage of comprehension of the natural language statement of the problem. Most of these researchers, however, stop short of defining how natural language is somehow "translated" into correct mathematical relationships or into a solution.

As researchers learn more about the processes of reading and the difficulties children have with mathematics, detailed analyses of the requirements of word problem solving are providing common ground for interdisciplinary results, including results on semantic types of problems and children's logico-mathematical competencies (Riley & Greeno, 1988), the locus of low-achieving students' eye-fixations (Hegarty, Mayer and Green, 1992) and the developmental sequence of central numerical structures (Case, 1985; Okamoto, 1992). Lacking in the above work, however, is an account of how the student's ability to arrive at mathematical connections might depend on identifiable ways in which natural language expresses or facilitates such connections. Recent results concerning the effects of small changes in problem wording (Cummins, 1991; Davis-Dorsey, Ross & Morrison, 1991; DeCorte, Verschaffel & DeWin, 1985), the role of integrated propositions in memory (Trabasso & Sperry, 1985 and others) and the importance of short-term memory as a bottleneck in the comprehension process (Fletcher, 1986 and others) provide new evidence that problem difficulty is determined by text comprehension factors as well as by semantic structure and the development of central numerical structures.

The information processing and learning assumptions in this research reflect the idea that different kinds of natural language used to convey a problem situation have different effects on the ability to conceptualize the problem in terms of correct mathematical relationships. Specific problem wording can (i) highlight certain mathematical relationships (e.g., direct references to previous sets facilitate the process of text integration), and/or (ii) lead to situational interpretations of events or quantities which are easier to retain in memory. Children's ability to follow-up on explicit set references (or infer such references) is a crucial step towards recognizing the conditions that make an arithmetic operation appropriate for a given situation.

There is as yet no detailed hypothesis of how a student might proceed from linguistic (mis)understanding to mathematical (mis)understanding. A precise model of this complex process, i.e., a computer model, starting with the initial presentation of the problem would be consistent with Goldin's (1992) call for a comprehensive model of mathematical problem understanding. In particular, a model which focuses on early stages of processing would force a consideration of critical features of the problem solving process that are often overlooked and could suggest new ways of isolating word problem solving competencies.

This paper describes our recent implementation of a computer model which simulates text integration processes involved in word problem understanding. The empirical results present new measures of text comprehension which determine why a particular word problem may be difficult, especially for young "bottom-up" readers. In this sense, the computer model proposes a new fine-grained classification of problems which vary according to both logical-mathematical problem solving processes *and* the often overlooked but critical processes involved in text comprehension.

EDUCE and SELAH

The current computer model comprehends arithmetic word problems written in English in a word-by-word, sentence-by-sentence fashion. The model is composed of two components: EDUCE and SELAH. EDUCE is an expectation-driven conceptual analyzer adapted from the work of Schank and Riesbeck (1981) and developed according to principles found to be distinct for word problem solving (LeBlanc and Russell, 1989). SELAH is a text integration component which accepts EDUCE's canonical representations of individual sentences and instantiates arithmetic actions based on explicit conceptual actions or direct or implicit references between sets (LeBlanc, 1991). In particular, the computer model keeps track of (i) the number of text integration inferences that are required and (ii) the load on working memory while performing text integration across sentences. It should be recognized that looking at the linguistic stage of word problem understanding is not just a matter of tacking on linguistic comprehension prior to mathematical considerations. We are assuming not that language and mathematics are separate components, but rather that they are intertwined, or that the mathematics is "embedded" in the natural language statement.

Experiments

Children's probability of solution as obtained in previous studies (Riley and Greeno, 1988) was analyzed in exploratory regression analyses as a function of 4 predictor variables, where the predictor variables were measures of text comprehension processes as simulated by the text integration component, SELAH. The word problems solved by the computer model are a classic benchmark set of 18 types of addition and subtraction problems. A sample problem is shown below:

> Kathy and Jacob have 8 soda cans altogether. Kathy has 5 soda cans. How many soda cans does Jacob have?

Procedure

The 18 problems were input to the model one at a time. For each problem, the model read and solved the problem while printing out a record of its text integration processes. The model's output for each problem was then analyzed to obtain scores for the 4 predictor variables to be used in the regression analysis. A summary of the 4 variables is described next in the context of the model's comprehension of the word problem shown above.

(1) AVGMEM. The average number of conceptual units in memory. This is the running sum of the number of conceptual units which appear in the model's working memory at the end of each sentence divided by the number of sentences in the problem. A conceptual unit is defined as either (i) a single conceptual set (e.g., [Kathy & Jacob's 8]); (ii) an arithmetic action (e.g., JOIN or SEPARATE_FROM); or (iii) an action and conceptual set(s) that have been "chunked" together. For example, in Table 1, the number of concepts in working memory at the end of the first sentence is two: ([K&J's 8] and JOIN). In the second sentence, the presence of the arithmetic action SEPARATE_FROM in conjunction with two conceptual sets with *known* quantities results in a chunking (or merge) of three concepts into two concepts. Specifically, the three concepts, "[Kathy's 5] [SEPARATED_FROM] [K&J's 8]," are chunked into two concepts: "[SEPARATE Kathy's 5] [FROM K&J's 8]." Two conceptual sets and an associated action are chunked only if *both* sets have *known* quantities. In the other case, where one of the sets has an *unknown* quantity, the three concepts [Kathy's *some*] [SEPARATED_FROM] [K&J's 8] are counted as three concepts, reflecting the model's sensitivity to a difference between remembering concrete vs. abstract quantities. Performing such chunking when *both* quantities are known reflects the fact that all the information needed to solve the problem (i.e., quantity-action-quantity) is known at this point. It may well be the case that these three concepts are actually chunked into one concept, however the more conservative chunking hypothesis (of three concepts into two concepts) is currently used throughout. In Table 1, the number of conceptual units in working memory at the end of each of the three sentences is 2, 2, and 3 respectively, for a total of 7. The average number of concepts in memory across the three sentences is thus $(2+2+3)/3 = 2.33$

(2) AVGINF. The average number of inferences made. This is the running sum of the number of inferences which are made divided by the number of sentences in the problem. An inference is defined here as: (i) establishing a relationship between two sets when that relationship is not described in the text or (ii) instantiating an arithmetic action (e.g., JOIN) when the text does not mention a significant action. Text integration and arithmetic action inferences are made when the conceptual representation of a new sentence (as produced by EDUCE) lacks the explicit information needed to establish text connections between conceptual sets or instantiate an appropriate arithmetic action. As shown in Table 1 in the first sentence, SELAH infers the JOIN action by noting (i) that the current set is partitioned by ownership and (ii) that the word "altogether" implies that this current set is the result of JOINing two (currently unknown) sets. In the second sentence, SELAH infers that the 5 cans that Kathy has are "*part of* the previously established set

Table 1 Computer model's processing summary for one of the eighteen problems.

Kathy and Jacob have 8 soda cans altogether. Kathy has 5 soda cans. How many soda cans does Jacob have?				
	Sentence			
	1	2	3	Average
Processing Summary	make set; infer JOIN	make set; infer 5 part-of 8; infer SEP-FROM; REACT	make set; infer ? part-of 8; infer ? is result of SEP-FROM	
Concepts in Memory	[K&J's 8] JOIN	[SEP K's 5] [FR K&J's 8]	[SEP K's 5] [FR K&J's 8] [RE J's ?]	
# of concepts	2	2	3	2.33
# of inferences	1	2	2	1.67

of 8." (An example sentence which would *not* have required this inference is: "Kathy has 5 *of them.*" Here the phrase "5 of them" would create a direct link to the previous "set of 8," so a text integration inference would not be required). SELAH makes another inference in the second sentence in order to build the set of [Kathy's 5]; since the five are now established as being *part of* the eight, the arithmetic action of SEPARATING-FROM is needed. The previously instantiated JOIN action is now suppressed in favor of SEPARATING-FROM. (This "reactivation" of an arithmetic action is noted in Table 1 as REACT; see variable #4 below). The average number of inferences across the three sentences is (1+2+2)/3 = 1.67

(3) MEM+INF. The sum of the variables AVGMEM and AVGINF, that is, the combination of the average number of concepts in working memory and the average number of inferences that are made. This variable reflects the theory that children's total processing capacity is made up of (i) what they must remember as well as (ii) what must be devoted to executing basic operations, such as making inferences (Case, 1982). In the context of arithmetic word problems, the fact that some problems require inferences may not be as critical as the number of concepts they must remember *while* they make those inferences.

(4) REACT. This variable indicates the situation where a previously instantiated arithmetic action must be suppressed in favor of a new action. A REACTivation (from JOIN to SEPARATE_FROM) is noted in sentence 2 of the example in Table 1.

Results

The four predictor variables were entered into a regression equation with the probability of solution for a specific grade level as the dependent variable. For each grade level, a stepwise regression was conducted.

Table 2 summarizes the results of the four regression analyses (grades K-3). For each grade, the variables listed in Column 2 are those that met a liberal .15 probability-to-enter criterion in the stepwise regression; they appear in the order that they were selected. Column 3 presents the proportion of total variance accounted for by the subset of variables in that row and all preceding rows *for each grade*. (The double line indicates a new grade-level and thus an independent regression analysis). For example, in grade K, MEM+INF accounts for .537 of the variance and MEM+INF and REACT together account for .690 of the variance. Columns 4 and 5 present the regression coefficients and standard errors for the selected variables. The last column presents the p values (two-tailed) associated with each of the regression coefficients. For example, in grade K, MEM+INF was significant at the .000 level and REACT was significant at the .016 level.

Discussion

An important result from this exploratory analysis is that MEM+INF is significant across all four grade levels (K-3). Because MEM+INF is always entered first in the stepwise analyses, the combination of "concepts to remember" and "inferences made" (MEM+INF) accounts for larger amounts of the variance in solution probability than either AVGMEM (the average number of concepts that must be held in working memory) and AVGINF (the average number of inferences that must be made in order to establish an integrated text).

Table 2 Results of regression analyses across four grade levels (K-3)

Grade	Variable	Variance Accounted For (R^2)	Regression Coefficient	Standard Error	p Value
K	MEM+INF	.537	-0.37	0.07	.000
	REACT	.690	0.32	0.12	.016
1	MEM+INF	.646	-0.47	0.07	.000
	REACT	.764	0.36	0.13	.015
2	MEM+INF	.721	-0.29	0.05	.000
3	MEM+INF	.577	-0.08	0.03	.019
	REACT	.664	-0.12	0.06	.066

One explanation is that a slight increase in either the number of concepts that must be remembered or the number of inferences required is not enough, in and of itself, to cause a working memory overload.

In addition, it is noteworthy that REACT appears as a significant predictor ($p < .05$) for only the youngest children: kindergarten and first grade. This confirms previous results which show that students often make wrong operation errors in line with an initial (although erroneous) activation of an arithmetic action (Cummins, Kintsch, Reusser and Weimer, 1988; DeCorte *et al.*, 1985). For example, in the word problem shown above, the most likely incorrect answer is thirteen (13), that is, students will add (i.e., JOIN) the two numbers in the problem rather than subtract 5 from 8. One explanation is that very young children do not have the ability to suppress previously activated information, in this case, a previously instantiated arithmetic action. In the current "bottom-up" implementation of the model, arithmetic actions can be (prematurely) instantiated by the presence of mathematical "keywords" (e.g., *altogether* means JOIN).

Implications for Teaching and Learning Mathematics

On the basis of the results from the regression experiments described above, predictor equations can be used to confirm previous research on the effects of problem wording and to extend the traditional measures of why one problem is easier or harder than another. The educational objectives are to increase teacher awareness of the multiple sources of problem difficulty and to show how slight changes in problem wording may affect children's solution success. The results of this analysis is a difficulty-differentiated network of problems that includes a multiple number of rewordings for each problem type in order to help teachers or a future computer-based learning environment determine a "next best" problem to present.

Table 3 Predicted and observed 1st grade solution probabilities for static, non-relational problem wordings

Problem Wording	Model's Prediction	DeCorte Data	Cummins Data
Traditional	0.47	0.43	0.30
altogether & of them	0.94	0.57	-
of them only	0.82	-	0.85

For example, the model confirms previous empirical results concerning children's success on the word problem discussed above, as well as on reworded versions of that same problem. This static (i.e., containing no significant actions), non-relational problem has received considerable attention in the literature, mostly due to its high level of difficulty. As discussed above, slight changes in wording (e.g., introducing the phrase "of them") may facilitate the processes of text integration and make this problem easier, as confirmed in studies with children (Cummins, 1991; DeCorte *et al.*, 1985) and highlighted by the computer model. A summary of probability of solution results are shown in Table 3. In line with the empirical studies, the model predicts that including the phrase "of them" will help most children (i.e., solution probability increases) and that removing the term "altogether" in the first sentence while also including the phrase "of them" can help considerably. The similarities and differences between the model's predictions and other empirical findings are beyond the scope of this paper. Our intent here is to show how the model's sensitivity to slight changes in wording is being used to extend the traditional classification of word problems. In addition to the rewordings discussed above, the model is sensitive to other slight changes in problem wording, including changes in the sequence of events and the use of significant action language to describe relational situations (LeBlanc, 1993).

Future Directions

The ability to explain problem difficulty rankings in terms of the processing of particular wordings brings us closer to a more comprehensive or global model of arithmetic word problem solving processes. The work presented above focuses on text integration processes and suggests that, even within this specific domain, the interaction of two tasks (e.g., concepts to remember and inferences made) may cause difficulty, where each individual task may not. Similarly, at a higher level, it seems clear that language comprehension processes and their relation to mathematical processes should be included in any model of arithmetic word problem solving, since consideration of mathematical or linguistic processes alone do not reflect all that is going on in the process of word problem solution.

A related reason for studying linguistic and text integration processes along with mathematical processes is the determination of possible relationships or distinctions between the various developmentally determined competences which underlie each process. From the perspective of the development of number concepts, for example, Case (1985) has postulated successive conceptual structurings, built up recursively from preceding structures, which are necessary for a child to understand the mathematics of arithmetic word problems with differing semantic structures. Some linguistic competences could apparently be related to some of these structurings. An understanding of nonquantitative verbal comparisons, for example, may be a prerequisite for an understanding of quantitative comparison and may therefore refer to the same conceptual structure. In the case of mathematically significant natural language, the determination of mathematical connections in EDUCE/SELAH in effect presupposes the mediation of elementary competences or structurings common to both linguistic and mathematical expression. Other aspects of natural language processing, (e.g., the additional memory load imposed by an "unnatural" or inconsistent ordering of sentences or sentence components) are presentation- rather than mathematics-related and may therefore require different competences, in addition to the ability to handle both textual and mathematical processing at the same time, as discussed above. The described research provides the future possibility to integrate competences for both tasks into a developmental theory which relates stages of mathematical understanding with stages of natural language understanding.

References

Case, R. (1985). *Intellectual development: Birth to adulthood.* New York: Academic Press.

Case, R. (1982). General developmental influences on the acquisition of elementary concepts and algorithms in arithmetic, In *Addition and subtraction: A cognitive perspective,* Carpenter, T.M., Moser, J.M., and Romberg, T.A. (Eds.). Hillsdale, NJ: Lawrence Erlbaum Associates, 156-170.

Cummins, D.D. (1991). Children's interpretations of arithmetic word problems.*Cognition and Instruction,* v8(3), 261-289.

Cummins, D.D., Kintsch, W., Reusser, K., and Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology,* v20, 405-438.

Davis-Dorsey, J., Ross, S.M., and Morrison, G.R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology,* v83, 61-68.

DeCorte, E., Verschaffel, L., and DeWin, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology,* v77, 460-470.

Fletcher, C.R. (1986). Strategies for the allocation of short-term memory during comprehension. *Journal of Memory and Language,* v25, 43-58.

Goldin, G. (1992). On developing a unified model for the psychology of mathematical learning and problem solving. *Proceedings of the 16th International Group for the Psychology of Mathematics Education,* v3, 235-262.

Hegarty, M., Mayer, R.E., and Green, C.E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology,* v84(1), 76-84.

LeBlanc, M.D. (1993). A computer model of the role of text integration in the solution of arithmetic word problems. Unpublished doctoral dissertation, Department of Computer Science, University of New Hampshire, Durham, NH.

LeBlanc, M.D. (1991). From natural language to mathematical representations: a model of "mathematical reading." *Intelligent Tutoring Media,* vol. 2, no. 3/4, 149-158.

LeBlanc, M.D. and Russell, S.W. (1989). Using conceptual information to help students obtain equations for algebra word problems. *Proceedings of the 6th Canadian Symposium on Instructional Technology.* Halifax, Nova Scotia, Canada, 191-194.

Okamoto, Y. (1992). The role of central conceptual structures in children's arithmetic word-problem solving. Paper presented at the Annual Meeting of the American Educational Research Association (AERA-92), San Francisco, CA, April.

Riley, M.S. and Greeno, J.B. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction,* v5(1), 49-101.

Schank, R.C. and C.K. Riesbeck (1981). *Inside Computer Understanding* (chapter 2). Hillsdale, NJ: Lawrence Erlbaum.

Trabasso, T. and Sperry, L.L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language,* v24, 595-611.