Impedance Effects of Visual and Spatial Content upon Language-to-Logic Translation Accuracy

Dave Barker-Plummer (dbp@stanford.edu)
CSLI, Stanford University
Stanford, California, 94305, USA

Robert Dale (Robert.Dale@mq.edu.au)
Centre for Language Technology, Macquarie University
Sydney, NSW 2109, Australia

Richard Cox (rcox@inf.ed.ac.uk)
School of Informatics, University of Edinburgh
Edinburgh, EH8 9AB, UK

Abstract

There is a body of work that suggests that those elements of the cognitive architecture responsible for processing, on the one hand, visual information (essentially visual properties of objects), and, on the other hand, spatial information (spatial relationships between objects), may compete with each other for resources. In this paper, we explore whether and to what degree the processing of visual and spatial information interferes with the task of translation from natural language into logic, a skill that students often find difficult to master. Using a large corpus of student data, we determine correlations between difficulty and the particular properties used in the sentences, with implications for pedagogical design.

Keywords: first-order logic, logic teaching and learning, visuospatial reasoning, visuospatial working memory, educational data mining, instructional design, visual impedance

Introduction

There is widespread agreement that the human visuospatial cognitive system consists of two dissociated but not entirely independent subsystems: one for processing visual information (e.g. object size, shape) and another for processing spatial information (e.g. the locations of objects with respect to each other). There is also the suggestion in the literature that these subsystems are used whether the information is perceived directly via external stimuli, or derived internally, via mental imagery (Logie, 1995; Baddeley, 2007). Further research suggests that particular combinations of visual and spatial information are more or less easily integrated during cognitive processing, as evidenced by impeded performance on reasoning and text comprehension tasks (Schuler, Scheiter & Gerjets, 2009). It seems that differences in the types of source information can lead to competition for cognitive resources in working memory, which in turn leads to poorer performance. These effects have been shown to occur whether information is presented as external stimuli, or as the result of mental imagery, and independently of the modality in which the information is presented.

Against this background, we investigate whether visual or spatial impedance effects are observed in a linguistically oriented cognitive task in which students translate natural language (NL) sentences into first-order logic (FOL). The ability to perform this task is a key skill in learning formal logic and related fields, such as mathematics, which require the formalization of informally presented information. The work described here is part of an ongoing project in which we investigate the factors which might make learning in these areas difficult (see e.g., (Barker-Plummer, Cox, Dale & Etchemendy (2008), Cox, Dale, Barker-Plummer & Etchemendy, (2008)).

Our study extends the field in two ways: by separately analyzing the effects of two kinds of visual information (object size and object shape) which are usually treated together, and by examining how those visual factors interact with each other and with spatial information.

Background: Human Visuospatial Processing

Baddeley's (2007) model of working memory contains a visuospatial sketchpad (VSSP), a phonological loop (for speech processing), plus an episodic buffer for holding and integrating diverse types of information. Attentional resources and rehearsal across the three working memory modalities is managed by a central executive. The VSSP partitions visuospatial working memory into two components: memory for spatial location and an object-based short-term memory. The VSSP is proposed as a storage system capable of integrating visual and spatial information (Baddeley, 2007; Baddeley & Hitch, 1974; Logie,1995; Logie & van der Meulen, 2009).

Recent evidence (Klauer & Zhao, 2004) provides support for dissociation between the visual and spatial subsystems and provides support for Logie's (1995) model of the VSSP. That model proposes a visual cache (for storing features such as shape, size, and colour) and an 'inner scribe' that deals with spatial and movement information.

Baddeley (2007) assumes that visuospatial information "may be encoded in the sketchpad either through perception, from long term memory (LTM), or via a combination of both" (p. 93). The VSSP, then, provides "a way of integrating visuospatial information from multiple sources" (p.101).

Short term memory for objects encodes features such as shape, size, orientation and texture. The visual system seems to readily combine and encode several features of any particular object (e.g.size and shape, shape and texture) as easily as a single feature of an object. The capacity limitation for objects in short-term memory (STM) seems to be more for the number of objects than for the number of features per object. Baddeley (2007) suggests that, for most people, the optimal number-of-objects versus number-of-features tradeoff in short-term memory seems to be 16 features distributed over 4 objects

Schuler et al. (2009) review research suggesting that written stimuli (text) may be processed in the VSSP in addition to auditory and phonological processing if it contains spatial information and/or information about visual features of objects (e.g. De Beni, Pazzaglia, Gyselink & Meneghetti, 2005).

Although the visual and spatial cognitive subsystems are dissociated to some extent, they are not completely independent, as evidenced by the fact that some combinations of information are more efficiently processed than others. Schuler et al. (2009) showed subjects coloured drawings of fish accompanied either by written spatial information (*The pectoral fin lies between the two dorsal fins*) or written visual information (*The pectoral fin has the same light brown color as the dorsal fins*). In one condition of the experiment, subjects were presented with the visual or spatial information aurally (spoken), and in another, subjects read the text. Learners given text with spatial content showed worse recall than those given visual text content, irrespective of presentation modality (written or spoken).

Knauff & Johnson-Laird (2002) used as stimuli relational terms of the form *The hat is above the cup, The cup is above the fork, Does it follow that the hat is above the fork?* The stimuli were varied according to information type. Examples of the kind just given are visuospatial (*above–below*), while pure-visual examples used terms such as *cleaner–dirtier* and pure-spatial examples used terms such as *north–south*. A control condition employed relations such as *better–worse*. It was found that visual relational terms were associated with longer reaction times (RT) in subjects' reasoning compared to control relations. Visuospatial relations produced faster RTs than control relations, with spatial-only relations producing the fastest RTs. There was no difference in error rates across the four conditions defined in terms of correct assessment of the truth of conclusions to valid and invalid inferences. Knauff & Johnson-Laird (2002) conclude that "...the principal effect is that visual relations slow down reasoning, relative to the other three relations" (p 368). Those authors termed the effect that they observed 'visual impedance', and concluded that irrelevant visual detail can impede reasoning. For Knauff & Johnson-Laird (2002), 'irrelevant visual detail' seems to means visual attributes of objects that do not assist the reasoner in building spatial mental models.

It is interesting to note that the Schuler et al. (2009) study demonstrated what might be termed 'spatial impedance', in contrast to Knauff & Johnson-Laird's (2002) reported visual impedance effect. However, the two studies differed substantially in at least two ways: first in terms of task (recall of information vs reasoning with information) and, secondly, in terms of cognitive source (perception vs mental imagery).

In this paper we explore the effects of visual and spatial text content in a different kind of task, that of translating natural language sentences into first-order logic. Our aim is to determine whether there is evidence of visual or spatial impedance effects in this domain.

Tasks of this kind have high ecological validity, since they arise in the sciences and in mathematics when information is translated into formal notations. The data we use in this study is collected from students participating in tasks designed for instruction of this key skill. A more detailed understanding of the cognitive processes in this task therefore has the potential to inform the design of instructional materials in these important subjects.

The Corpus

In order to investigate the presence of visual and spatial impedance effects, we data-mined a large-scale educational corpus in the area of logic teaching (Barker-Plummer, Dale, & Cox, 2011). The corpus consists of student-generated solutions to exercises in *Language, Proof and Logic* (LPL; Barwise, Etchemendy, Allwein, Barker-Plummer & Liu, 1999), a courseware package consisting of a textbook together with desktop applications which students use to complete exercises.[1] The book offers an introductory course in formal logic for early undergraduates. Students may submit answers to 489 of LPL's 748 exercises[2] to The Grade Grinder (GG), a robust automated assessment system that has assessed more than 2 million submissions of work by more than 55,000 individual students in the period 2001–10; this population is drawn from approximately a hundred institutions in more than a dozen countries.

One type of exercise in LPL involves translating natural language (NL) sentences into first-order logic (FOL). The corpus contains a total of 275 sentences which students are asked to translate and submit to the Grade Grinder. Most of these sentences refer to a blocks world of objects arrayed on a checkerboard. The objects may have visual properties such as shape (cubes, tetrahedra, dodecahedra) and/or size (small, medium, large). They may also have spatial relationships with other objects on the grid (in front of, between). The remaining sentences involve completely different vocabularies involving either numbers, or people and their pets. The Grade Grinder considers a translation for a sentence to be correct if it is provably equivalent to a reference solution.[3]

From the 275 sentences we identified a subset of 129 sentences for study, omitting those that included references to temporal information and other semantic phenomena that were idiosyncratic with respect to our current investigation. In contrast with earlier studies, we investigate the effect of size and shape information separately. While these are both types of visual information, shape information is discrete, while size information is generally considered scalar. So we partitioned these sentences into eight categories according to whether size, shape and spatial information are present in the sentence (Figure 1). In that figure we refer to the type of sentence in row 6 as 101, meaning that the sentences in this class

[1] See `http://lpl.stanford.edu`.

[2] The other exercises require that students submit their answers on paper to their instructors.

[3] There are infinitely many correct answers for any sentence, so a theorem prover is employed to determine equivalence.

Type	*n*	Example NL sentence in category
000	10	*Max is a student, not a pet*
001	15	*At least one of A, C, and E is a cube*
010	15	*B is larger than both A and E*
011	25	*Everything smaller than A is a cube*
100	18	*C is in back of A but in front of E*
101	24	*B is not to the left of a cube*
110	6	*Either E is not large or it is in back of A*
111	16	*B is to the right of a large cube*

Figure 1: Examples of sentences in each of the eight categories. *Type* indicates the presence or absence of spatial, size and shape information respectively (so, 111 means all three are present). *n* refers to the number of sentences in the category.

Type	pincorr	SD	Mean no. trans.	SD
000	.25	.15	6226.80	4902.77
001	.09	.07	13173.87	10944.13
010	.15	.19	18488.27	9578.00
011	.26	.21	10680.08	6277.08
100	.19	.22	19009.44	7915.24
101	.23	.15	9689.00	6232.25
110	.29	.24	15255.33	8883.13
111	.20	.14	10170.25	8627.53

Figure 2: Proportion of submitted student translations that were incorrect (pincorr) for each category of sentence with standard deviations (SD), together with the average number of translations considered and standard deviations.

contain spatial information (indicated by the leading digit) and shape information (indicated by the last digit). In the text we refer to this class as **space+shape** sentences.

Method

Measuring Problem Difficulty

Our measure of the difficulty of the translation task for a particular sentence is the proportion of the initial attempts at translation that are in error, which we label **pincorr**. Pincorr values range from 0–1, with smaller values indicating fewer errors. Figure 2 shows the pincorr values for each of the sets of sentences, and the average (mean, standard deviation) number of subjects contributing to these values.[4]

Note that pincorr is the proportion of *initial* attempts by a subject that are in error. The Grade Grinder places no limit on the number of times that an exercise may be attempted, and the corpus contains many attempts by the same subject at translating the same sentence. These translation attempts are presumably revised on the basis of GG's feedback on earlier attempts, so we calculated the translation error rates by considering only the initial submission of a sentence by an individual student.

The pincorr value for the class with no size, shape and spatial information (row 1 of the table) must be treated with some caution. The LPL package includes desktop software which enables students to build 'worlds' in which they can evaluate the truth of their sentences, providing a way for students to test their solution.[5] Students can determine whether a translation is *incorrect* if they can build a world in which the NL sentence has a different truth-value from the candidate FOL translation, but they cannot definitively check whether a solution is correct. Sentences which contain no size, shape or spatial information cannot be checked for error in this way, and we would therefore expect the error rate for these sentences to be higher than that of the other, testable, classes. This class is also notable for the relatively small number of subjects translating these sentences.

Given the different levels of error across the eight sentence types shown in Figure 2, we examined the effect of membership in each of the eight sets using pincorr as a dependent variable. Of course, it may not be the particular combinations of visual and spatial information that result in different levels of difficulty; other possibilities, for example, are the readability and informational complexity of the sentences. Below, we describe two separate analyses which control for these possibilities using covariates in our analyses. In the first we control for differences in readability across sentence classes using the Flesch readability index (Talburt, 1986) as a surrogate for sentence comprehension difficulty. In the second we use the presence of a binary predicate in the sentence as a surrogate for informational complexity.

Analyses

Readability as a covariate

A three-way analysis of covariance (ANCOVA) was performed. Each of the three factors (size, shape and spatial information) had two levels (present/absent). This analysis used the Flesch readability index as a covariate in order to control for the readability of the sentences. The interaction plot is represented by both graphs in Figure 3.

In order to further elucidate the components of the interaction, the three-way ANCOVA was partitioned into two separate, two-factor ANCOVAs.

The first analysis was conducted upon sentences that do not contain spatial information (i.e. the first four rows of Figure 1) and the second on those that contain spatial information (i.e. the lower four rows of Figure 1).

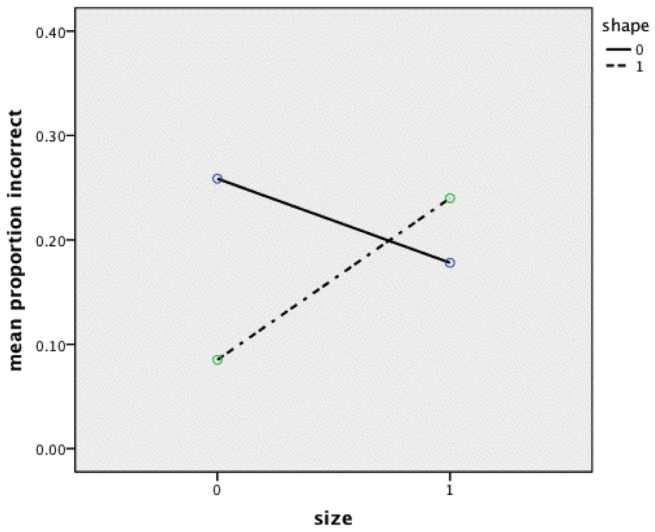
The two-way interaction graph for non-spatial sentences is shown in the upper graph of Figure 3.

Informational complexity as a covariate

Spatial information in the LPL blocks language concerns relations between objects. For example, one object may be *to the left of*, or it may *adjoin*, another object. The spatial language also contains one ternary relation, *between*. By contrast, visual information in the language predominately concerns the properties of objects. An object may be a *cube* (shape) or *small* (size). Relations, such as *smaller* (size) and *same shape* (shape), do occur in the language. However, the spatial fragment of the language is exclusively relational, and this offers

[4]The tasks completed vary by subjects represented in the corpus.

[5]These worlds are displayed in a graphical modality.



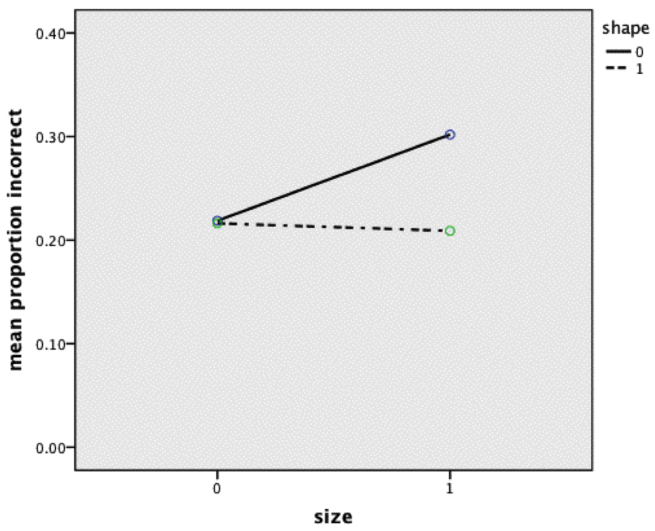
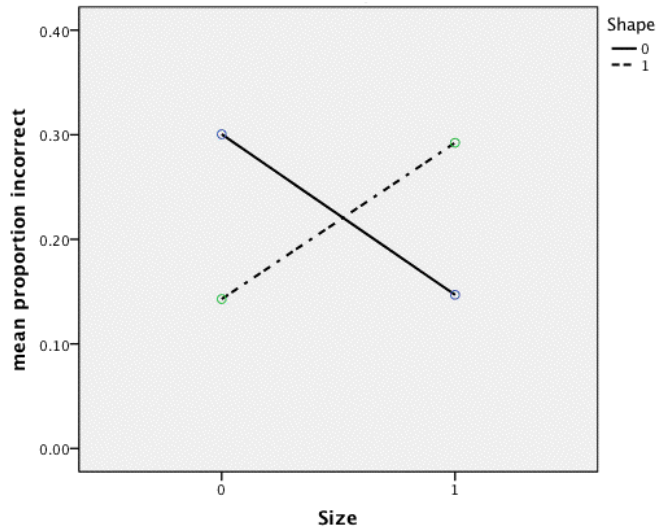


Figure 3: ANCOVA 3-way interaction plot showing sentences referring only to object shape and size (upper figure) and sentences referring to spatial location of objects as well as shape and size (lower figure). The Flesch index of readability was included as a covariate in the analysis; the plotted values are adjusted means and differ from those in Figure 2.



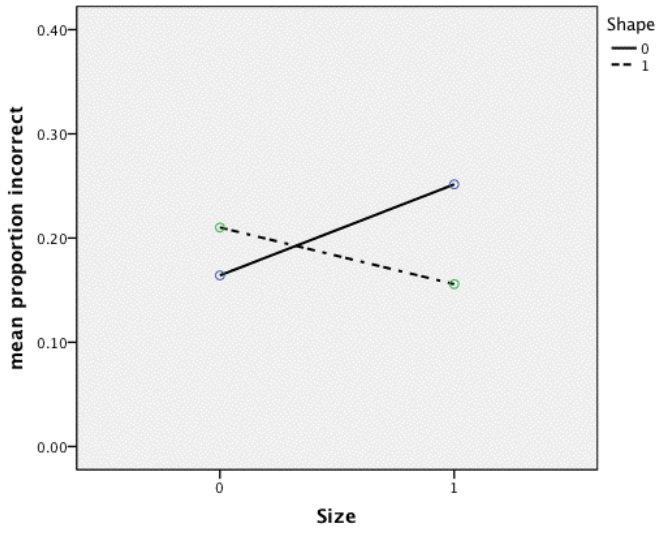


Figure 4: ANCOVA 3-way interaction plot showing sentences referring only to object shape and size (upper figure) and sentences referring to spatial location of objects and shape and size (lower figure). For each sentence in each group the presence of higher arity (binary) predicates was included as a categorical covariate.

an alternative explanation of variance in difficulty between spatial and visual sentences.

The distinction between the predominantly relational and predominantly property fragments of the language is a difference of informational complexity. We use the presence of a binary predicate as a proxy for this difference. Sentences containing one ore more binary relations (pincorr: $M = .235$, $SD = .211$) are significantly more difficult to translate ($t = -2.36$, $p = .02$) than sentences that contain no binary predicates (pincorr: $M = .166$, $SD = .118$), and the eight groups differ in terms of the number of binary-predicate-containing sentences they include.

In order to determine whether the effect that we observed was due to increased complexity of FOL sentence rather than visual/spatial impedance, we repeated the analyses using a binary covariate indicating, for each sentence in every group, whether or not it contains a binary predicate.

A three-way ANCOVA analysis was performed. The three-way interaction plot is represented by both graphs of Figure 4. The three-way analysis was partitioned into two two-way analyses. For non-spatial sentences, the two-way interaction graph is shown as the upper graph in Figure 4.

Type	pincorr (Flesch)	pincorr (Arity)
000	.26	.30
001	.08	.14
010	.18	.15
011	.24	.29
100	.22	.16
101	.22	.21
110	.30	.25
111	.21	.16

Figure 5: Summary adjusted pincorr 3-way interaction means, for each covariate.

A second two-factor ANCOVA was performed on sentences that contained size and/or shape information together with spatial information (the lower four rows of Figure 1).

Results

The results of both analyses, controlling readability and informational complexity, agree on the following:

- The three-way ANCOVAs interaction effects were significant (Flesch: $F(1,120)=5.51$, $p = .02$, Arity: $F(1,120) = 10.72$, $p = .001$). This indicates that the effects of size and shape information upon translation difficulty differ at different levels of the spatial factor.

- The two-way ANCOVAs with shape and size as independent variables for sentences with no spatial information (upper four rows of Figure 1, upper graphs in Figure 3 and Figure 4) revealed no main effect of size or shape, but a significant size-by-shape interaction (Flesch: $F(1,60) = 6.66$, $p = .012$, Arity: $F(1,60) = 13.4$, $p = .001$).

- The two-way ANCOVAs with shape and size as independent variables for sentences with spatial information (lower four rows of Figure 1, lower graphs in Figure 3 and Figure 4) showed no significant main effects or interactions.

Figure 5 shows the pincorr values from the ANCOVAs with each of the covariates. Considering these adjusted pincorr means, and writing '<' to mean 'easier to translate', we can sum up the trends as follows:

- Both of the analyses show that sentences involving only one type of information have lower values than sentences involving any combination of information types. Among these *homogeneous* sentences, **shape < size < space**.

- Sentences involving all three information types[6] are shown by both analyses to be the next hardest sentences to translate, *i.e.* they are more difficult to translate than homogeneous sentences, but easier than any pairwise combinations.

- In the first analysis, with Flesch as a covariate, the translation difficulty ordering is **space+shape < size+shape < space+size**. In the second analysis with arity as the covariate, the relative difficulty of **size+shape** and **space+size** are switched.

A striking effect was that, whereas sentences containing references to shape but not to either size or spatial information were the least error-prone to translate (M=.09, SD=.07, row 2 in Figure 2), when spatial information is added, the combination of spatial and shape information (M=.23, SD=.15, row 6 in Figure 2) significantly increases difficulty (t = -3.42, p = .002). In sentences without spatial information, combining size information with shape information (M=.26, SD=.21, row 4 in Figure 2) significantly increases difficulty (t = -3.12, p = .003) compared to only shape information (M=.09, SD=.07, row 2 in Figure 2).

Discussion

The interaction of visual and spatial features of sentences affects sentence translation difficulty with effects that are similar when controlling for each of two potential confounding factors, readability and informational complexity.

Taken together, the results suggest that the easiest-to-translate sentence types are those that contain just one visual or spatial type of information, with a relative difficulty of **shape < size < space**.

Contrary to Schuler et al. (2009) we did not find a *simple* negative effect of combining spatial information with visual information in a sentence. Rather, the type of visual information seems to make a difference: our results suggest that spatial information plus size information tends to produce more difficult-to-translate sentences than spatial information combined with shape information.

This suggests that research on visuospatial reasoning and visuospatial working memory needs to distinguish between *subtypes* of visual information. Visual features such as size, shape and perhaps color, may differ in terms of the demands they place (singly and in combination) upon working memory.

A surprising finding is that the **size+spatial** and **shape+spatial** classes both have higher pincorr values than that for the **size+shape+space** class. This result challenges theories which suggest that impedance effects result from competition for cognitive resources, since this would suggest that impedance effects observed in sentences containing two types of information should not be reduced by the addition of a third type of information (or of more visual information if space and shape are to be considered as one type).

Our study is closest in kind to that reported in Knauff & Johnson-Laird (2002). In both studies, information is presented in the form of sentences to be read, and these sentences contain different information types. However, our tasks vary in the number of types of information to be processed, in contrast to the tasks in Knauff & Johnson-Laird, which are each homogeneous. The implications of our findings for Knauff &

[6] We consider visual(size) and visual(shape) to be different information types. They are subcategories of visual information which differ, *inter alia*, in terms of whether they are discrete properties (shape) or scalar (size).

Johnson-Laird's (2002) 'visual-imagery impedance' hypothesis are not clear. In particular, their hypothesis makes use of the notion of 'irrelevant visual detail', referring to those visual attributes of objects that do not assist with the building of spatial mental models. In our task, shape, size or spatial information about objects—if mentioned in a sentence—is always relevant to the task of NL to FOL translation.

The results have implications for logic teaching. Instructors, when creating sentences for NL to FOL translation exercises designed to teach logical connectives, quantifiers, and concepts like implicature, will be better equipped for the principled design of learning exercises. They could, for example, consider introducing sentences that refer only to object shape, then later challenge students with sentences that describe objects in terms of, say, spatial position and size, at a stage when the student is more practiced and confident.

In further work we propose to address individual differences in cognitive processing of various forms of information. In earlier work we have demonstrated individual differences between students in multimodal (graphical and sentential) logic learning contexts (e.g. Stenning, Cox, & Oberlander, 1995). Students' analytical reasoning performance on constraint-satisfaction problems was shown to predict their propensity to develop flat versus 'nested' (broken-into-cases) styles of logical proof. Stenning et al. (1995) concluded that 'verbaliser/visualiser' conceptions of learning style are too simplistic: rather than preferring visual or verbal reasoning contexts, Stenning et al., (1995) found that students differed more in their tendency to stay in one modality (graphical or sentential) as opposed to switching between modalities. More recently, Blazhenkova & Kozhevnikov (2009) have proposed a three-dimensional cognitive style model in which people are held to differ in their learning style preferences for material containing object imagery, spatial imagery and verbal content. Exploiting the very large number of student submissions in our corpus, we plan next to study whether we can identify sub-groups of students who respond exceptionally to particular information-type configurations. The aim is to statistically identify such student clusters and to establish which current individual difference theory the cluster patterns support.

References

Baddeley, A. & Hitch, G.J. (1974) Working memory. In G. Bower (Ed.), *The psychology of learning and motivation*, (Volume 8, pp. 47–90). New York: Academic Press.

Baddeley, A. (2007), *Working memory, thought, and action*., Oxford University Press.

Barker-Plummer, D., Cox, R., Dale, R. & Etchemendy, J. (2008) *An Empirical Study of Errors in Translating Natural Language into Logic*. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (CogSci 2008).

Barker-Plummer, D., Dale, R., & Cox, R. (2011) The Grade Grinder Corpus Release 1.0: Student Translations of Natural Language into Logic. CSLI Technical Report, Stanford University.

Barwise, J., Etchemendy, J., Allwein, G., Barker-Plummer, D. & Liu, A. (1999) *Language, Proof and Logic*. CSLI Publications and University of Chicago Press.

Blazhenkova, O. & Kozhevnikov, M. (2009) The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, **23**, 638-663.

Cox, R., Dale, R. Etchemendy, J., & Barker-Plummer, D., (2008) Graphical Revelations: Comparing student Translation Errors in Graphics and Logic. In J. Howse, J. Lee and G. Stapleton (eds), *Proceedings of the 5th International Conference on the Theory and Application of Diagrams* (Diagrams 2008). Springer.

De Beni, R., Pazzaglia, F., Gyselink, V. & Meneghetti, C. (2005) Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology*, **17(1)**, 77–95.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Harvard University Press, Cambridge, MA.

Kintsch, W. (1991) A theory of discourse comprehension: Implications for a tutor for word algebra problems. In M. Carretero, M. Pope, R-J. Simons & J.I. Pozo, *Learning and instruction: Volume 3*, A publication of the European Association for Research on Learning & Instruction, Oxford: Pergamon Press.

Klauer, K.C. & Zhao, Z. (2004) Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General*, **133**, 355–381.

Knauff, M. & Johnson-Laird, P.N. (2002) Visual imagery can impede reasoning. *Memory & Cognition*, **30(3)**, 363–371.

Logie, R.H (1995) *Visuo-spatial working memory*, Hove, UK: Lawrence Erlbaum Associates.

Logie, R.H. & van der Meulen, M. (2009) Fragmenting and integrating working memory. In J.R. Brockmole (Ed.) *The visual world in memory*, Hove, UK: Psychology Press, Taylor & Francis Group.

Schuler, A., Scheiter, K., & Gerjets, P. (2009) Does spatial verbal information interfere with picture processing in working memory? The role of the visuo-spatial sketchpad in multimedia learning. In S. Ohlsson & R. Catrambone (Eds.) *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp 2828–2833), Cognitive Science Society.

Stenning, K., Cox, R. & Oberlander, J. (1995) Contrasting the cognitive effects of graphical and sentential logic teaching: Reasoning, representation and individual differences. *Language and Cognitive Processes*, **10(3/4)**, 333 – 354.

Talburt, J., (1986) The Flesch Index: An Easily Programmable Readability Analysis Algorithm. *Proceedings of the 4th Annual International Conference on Systems Documentation*. (pp. 114-122). New York, NY: ACM.