Knowledge Structures and Their Application in CALL Systems

Extended abstract

Ivo Düntsch*
Faculty of Informatics
University of Ulster
Newtownabbey, BT 37 0QB,
N.Ireland
I.Duentsch@ulst.ac.uk

Günther Gediga*
FB Psychologie / Methodenlehre
Universität Osnabrück
49069 Osnabrück,
Germany
ggediga@luce.psycho.Uni-Osnabrueck.DE

Abstract

The paper describes two approaches towards knowledge assessment which can be applied within any CALL application. The first approach uses test items to define a universe of empirical knowledge states. We show that the relations between items ('are about equal', 'is harder than') and subjects ('are about equal', 'is better than') can be easily computed within a CALL system, and that the empirical results give insight into the structure of problems and their solutions. Given a standardized test procedure, a CALL system can use this method to describe the knowledge state of a tested subject either using a reference population, a knowledge structure generated by experts (see below), or by computing a knowledge structure using the given sample.

The second approach is based on querying four experts about the skills minimally necessary to solve problems of the item set of the first approach. The results of these queries can be transformed into theoretical knowledge structures, and the empirical data can be used to test the prediction of the experts. The results show that

- The experts disagree remarkably,
- Most of the experts did not express even simple relationships which are observable by the empirical data analysis,
- Situational factors such as the appearance of learning tasks in the same lecture are observable in the empirical data as well as in the opinions of the experts.

*Equal authorship implied

1 Structural Background and Tools

We consider a set U of subjects and a set P of problems as well as a relation $R \subseteq U \times P$ such that

(1) $\langle s, p \rangle \in R$ if and only if subject s solves problem p.

The set $K_s := \{p \in P : \langle s, p \rangle \in R\}$ is called the *(knowledge) state of subject s*, and the collection $\mathcal{K} := \{K_s : s \in U\} \cup \{\emptyset, P\}$ is called a *knowledge structure* . Each knowledge structure $\mathcal{K}$ is partially ordered by set inclusion $\subseteq$. We can say that a problem p is a *prerequisite* of problem q if every state which contains q also contains p; in this sense, we also call q *harder* than p. In other words, q is harder than p, if there is no subject which solves problem q but fails to solve problem p. Similarly, we can say that subject x is *better* than subject y, if $S_y \subsetneq S_x$.

A consolidated line diagram of these two partial orders such as Figures 1 or 2 can be obtained via the *formal concept analysis* approach presented in Wille (1982).

Suppose that S is a set of skills. A function $\gamma : P \rightarrow 2^{2^S}$ is called a *skill function* , if $\gamma(q)$ is a nonempty set of nonempty, pairwise incomparable subsets of S for each $q \in P$. We interpret the elements of $\gamma(q)$ as exactly those sets of skills which are minimally sufficient to solve problem q; in other words, each $X \in \gamma(q)$ is a set of skills sufficient to solve q, while each proper subset of X is not. If $\gamma(q)$ contains only one element X, the expert states that each strategy to solve problem q must contain the skills which are specified by X. If $\gamma(q)$ contains n subsets of S, there are n essentially different strategies to solve problem q. In Düntsch & Gediga (1995) we have shown that each skill function uniquely determines a knowledge structure on P. In other words, an expert who specifies a skill function at the same time builds a knowledge structure.

Knowledge assessment via knowledge structures and skill functions has been used successfully in other areas (see Falmagne et al. (1990); Düntsch & Gediga (1995)), and below we describe an application within the context of a language learning environment.

2 Empirical Knowledge Structures

2.1 The Test Data

The test data were obtained in a *German as a foreign language* (GFL) course which is documented in more detail in Fox & Hamilton (1996).

The course was at beginner-level, and introduced learners to idioms and vocabulary of introductions (greetings,introductions, origins, meeting and contacting people), basic structures (questions and answers), people's identity, expressing wishes, role-play situations (pubs, making telephone calls), basic grammar (present tense, conjugations), basic sentence structure (SV etc.), identifying places and things, counting, language functions (suggesting, advising etc.), communicative tasks, SVO, indefinite and definite, negative and affirmative.

In class as well as by a CALL system, a range of exercises was given which aimed at developing and enforcing these skills. At the start of the course, 18 subjects were assigned to a CT-group (Computer+Tutor; group code 1) and eight to a T-group (Tutor only; group code 2); seven additional subjects were asked to join the CT-group in a later stage of the course (group code 3).

2.2 Data Analysis

The test was given after 6 hours of lectures, and contained 18 questions listed in Table 1.

We have constructed the knowledge structures belonging to the respective groups (T, CT, T $\cup$ CT), as well as the partial orders on the subjects and problems as described in Section 1. The transitive line diagram based on the generated concept lattice of the combined set is given in Figure 1.

Although the line diagram looks rather complex, we give it as an example of the method; in Fig. 2 we present a reduced version. The diagram should be read as follows:

- If we compare subjects an ascending line from subject x to subject y means *y is better than x,* since y was able to solve each problem which x could solve, as well as (at least) one additional problem; for example subject 112 (on the top right) is better than subject 114.
- Ascending lines between item nodes should be read as "is harder than", e.g. A_{18} (on the top left) is harder than A_{17}, because there was no subject who did not solve A_{17}, but did solve A_{18}.
- An ascending line from a subject x to an item q indicates that x was not able to solve q; for example, subject 112 was not able to solve problem A_5.
- An ascending line from an item q to a subject x to indicates that x was able to solve q; for example, subject 333 was able to solve problem A_4.

Table 1: Test questions

A_1	Wie . . . es Ihnen? (a) trinkt (b) ist (c) kommt (d) geht
A_2	Wo . . . Sie? (a) kommen (b) nehmen (c) wohnen (d) heissen
A_3	. . . kommen Sie? (a) wo (b) wer (c) wie (d) woher
A_4	Verzeihung, wie . . . Ihr Name? (a) bist (b) ist (c) sein (d) sind
A_5	Wie . . . das auf deutsch? (a) heisse (b) heissen (c) heisst (d) heiss
A_6	. . . du auch Englisch? (a) sprichst (b) spricht (c) spreche (d) sprechen
A_7	Siebzehn (a) 27 (b) 6 (c) 17 (d) 16
A_8	Dreiundzwanzig (a) 30 (b) 32 (c) 20 (d) 23
A_9	Guten M. . .
A_{10}	Verzeihung, w. . . ist Ihr Name?
A_{11}	Wie s. . . man das?
A_{12}	Wo w. . . Sie?
A_{13}	Konjugieren Sie bitte "sein".
A_{14}	Konjugieren Sie bitte "wohnen".
A_{15}	Antworten Sie oder fragen Sie: Das ist Frau Lim. . . . ?
A_{16}	Antworten Sie oder fragen Sie: Sie kommt aus Korea. . . . ?
A_{17}	Antworten Sie oder fragen Sie: Wie heissen Sie? . . .
A_{18}	Antworten Sie oder fragen Sie: Wo arbeiten Sie? . . .

Scanning the line diagram, a few items of relational information can be obtained, and the results are given in Table 2.

Although this type of data representation is rather "soft" (because it uses the relational information of the data as they are), we can draw some conclusion from the structure:

- Problem A_9 is at the bottom of the structure and it was solved by everyone; therefore, A_9 contains no information to separate the subjects.

Figure 1: The empirical lattice

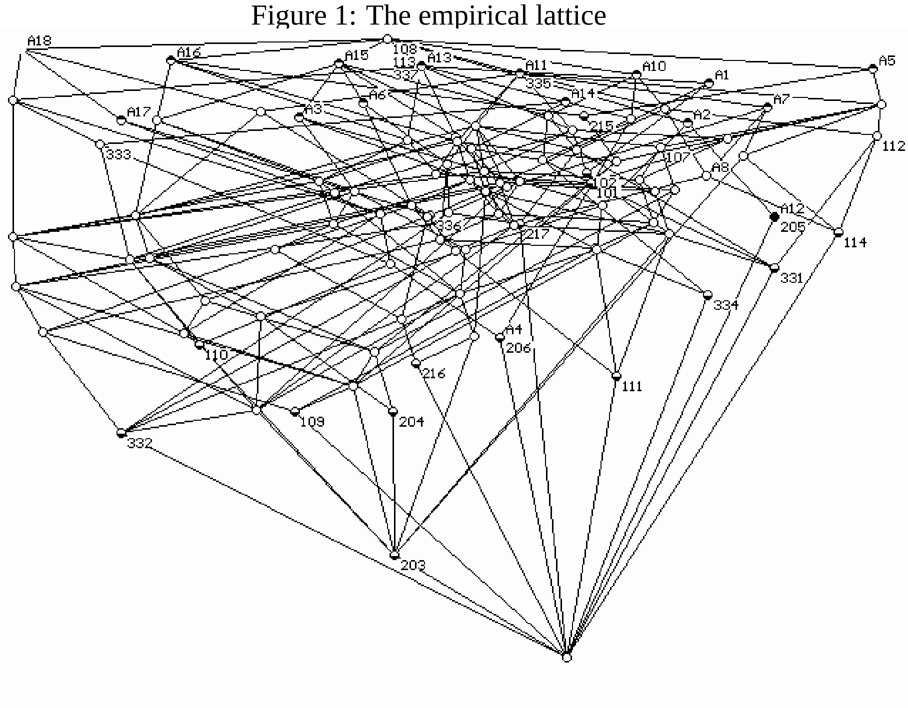


Table 2: Lattice structure excerpt

Pairwise incomparable items:		$A_2, A_5, A_{10}, A_{11} A_{13}, A_{14}, A_{15}, A_{16}, A_{18}$	
Some comparable items			
A_1:	$A_9 < A_1 < A_{11}$	A_6:	$A_9 < A_6 < A_{15}$
A_3:	$A_9 < A_3 < A_{15}$	A_7:	$A_9 < A_8 < A_7 < A_{11}$
A_4:	$A_9 < A_4 < A_{10}$	A_8:	$A_9 < A_8 < A_7 < A_{11}$
	$A_9 < A_4 < A_{13}$		$A_9 < A_8 < A_{10}$
	$A_9 < A_4 < A_{14}$	A_9:	Smallest element
	$A_9 < A_4 < A_{17}$	A_{12}:	$A_9 < A_{12} < A_2$
	$A_9 < A_4 < A_{18}$	A_{17}:	$A_9 < A_{17} < A_{18}$

- The relations $A_4 < A_{10}$ and $A_{12} < A_2$ should be included in any error theory which describes the test results, because the phrases used are identical, but the harder items contain less information how to solve the item.

- A_4 is a key item of the test, because not solving A_4 implies not solving A_{10}, A_{13}, A_{14}, and A_{18}.

It was noted earlier (Fox & Hamilton (1996)) that the subjects of the T-group were not as successful as the subjects of the CT groups. This can be read off the Figure by noting that no subjects of the T group are located near the top node, whereas some of the CT-group can be found there. The low scoring subjects of both groups can be described differently: The set of low scoring subjects $L = \{203, 204, 206, 216\}$ of the T- group have problems to pose questions, because any item of the set $\{A_2, A_3, A_{15}, A_{16}, A_{17}, A_{18}\}$ dominates at least 3 elements of set L. The critical items of the low scoring subjects in the CT-groups are $\{A_3, A_6, A_{11}, A_{13}, A_{14}, A_{15}\}$, which indicates that the conjugation of verbs is a problem in that group.

2.3 Reducing the Lattice Complexity

The lattice representation (Fig. 1) of the data is very complex. This fact indicates that most of the subjects are not comparable using the full set of items to represent the knowledge of the subjects. Because the study was aiming at the determination of a difference between the T-group and the CT-group in terms of their knowledge, we have developed a procedure to reduce the lattice complexity by disregarding items which cause incomparability between both subject groups.

More formally, we define for each $\emptyset \neq Q \subseteq P$ the relation $R_Q \subseteq U \times Q$ by

(2) $\langle x, q \rangle \in R_Q$ if and only if x solves problem q.

Observe that R_Q is just the relation R of (1) with its range restricted to Q, so that in particular, $R_U = R$. We also set

$$\mathrm{ran}_Q\, x := \{q \in Q : \langle x, q \rangle \in R_Q\}.$$

For each $\emptyset \neq Q \subseteq P$ we now define the following relations on $CT \times T$:

$$\begin{aligned} x <_Q y &\overset{\text{def}}{\iff} \mathrm{ran}_Q\, x \subsetneq \mathrm{ran}_Q\, y, \\ x >_Q y &\overset{\text{def}}{\iff} \mathrm{ran}_Q\, x \supsetneq \mathrm{ran}_Q\, y, \\ x \equiv_Q y &\overset{\text{def}}{\iff} \mathrm{ran}_Q\, x = \mathrm{ran}_Q\, y, \\ x \#_Q y &\overset{\text{def}}{\iff} \text{ none of the above.} \end{aligned}$$

Finally, we let $O : 2^U \setminus \{\emptyset\} \rightarrow \mathbb{N}$ be defined by

$$O(Q) = \mathrm{card}(\equiv_Q) + \mathrm{card}(\#_Q).$$

In minimizing the objective function O, we obtain a knowledge space with reduced complexity (omitting the items A5, A6, A8, A10, A11, A16), which is presented in Fig. 2.

Members of the CT-group are coded by "1", and members of the T-group are coded by "2". It is obvious that the members of the T-group have much more difficulties with the chosen items than the subjects within the CT-group. In order to test this differences, we used a random labeling approach using the statistic

$$g(Q) = \frac{\mathrm{card}(>_Q) - \mathrm{card}(<_Q)}{\mathrm{card}(>_Q) + \mathrm{card}(<_Q)},$$

and found out that the CT-group dominates the T-group (α=0.05). On the contrary, the simple scale sum approach using the same data (Fox & Hamilton (1996)) shows no significant differences between both groups. It should be noted that the optimization procedure did not optimize the differences between the groups (if so, the used randomization procedure would be invalid!), but minimized the number of not comparable elements, which is simply a reduction of error variance.

3 Knowledge Structures Generated by Experts

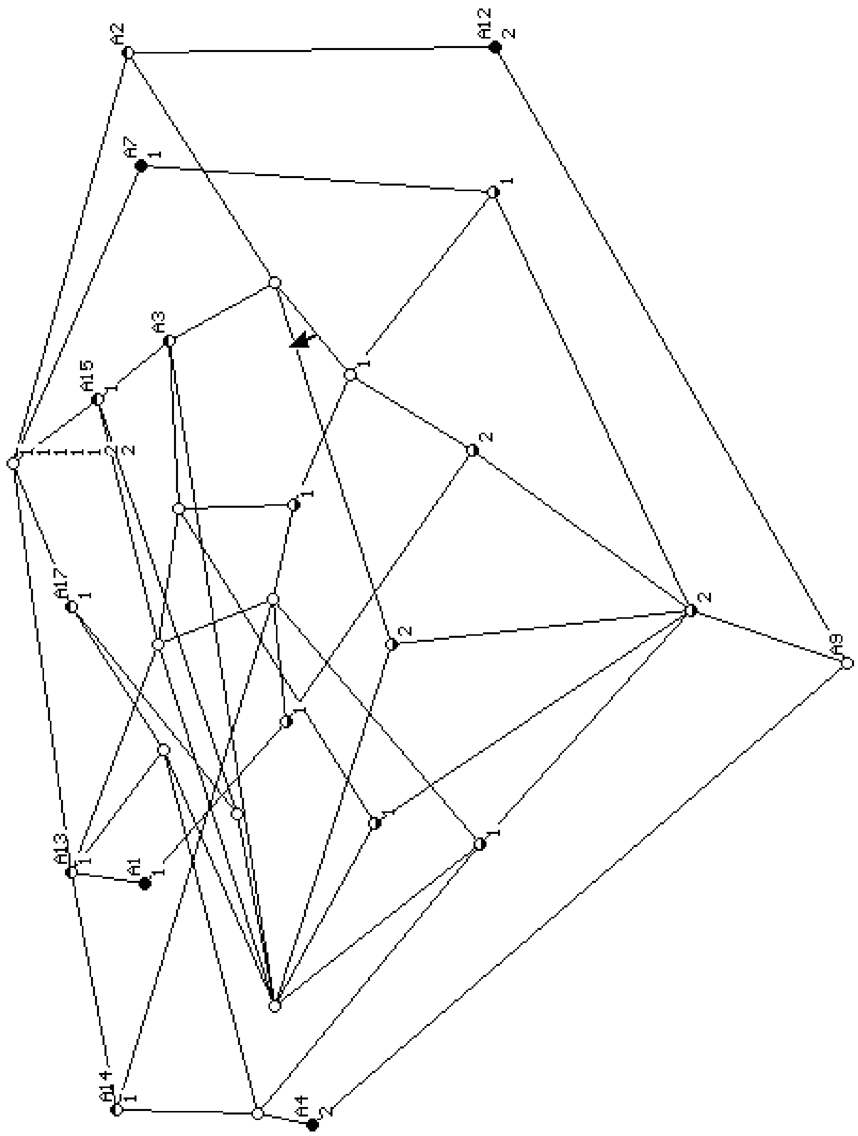
3.1 The Data Base

In order to construct the theoretical counterpart of the empirical knowledge structure generated in the preceding section, we asked four experts to perform an expert analysis of the test (see Koppen & Doignon (1990); Düntsch & Gediga (1996) for the details): Given a test item, every expert had the task to specify the minimal combination of language skills a subject should have in order to be able to solve the item. Since there are sometimes several ways to solve a problem, the expert was given the chance to define up to three ways to solve the problem. In Düntsch & Gediga (1995) we have shown that the skill function approach enables the expert to describe any possible knowledge structure.

Four experts performed the procedure:

- Expert 1 is a male native English language tutor, teaching GFL.
- Expert 2 is a female native German language teacher, teaching EFL in Germany.

Figure 2: The reduced empirical lattice



- Expert 3 is a male native American language teacher, teaching EFL in Germany.
- Expert 4 is a female native English language teacher, teaching GFL who

constructed the achievement test.

3.2 Data Analysis

Expert 1 and Expert 2 orient their assignment of skills at the lecture position of the learning material, assuming that one basic skill learned within a lecture is sufficient to predict the outcome within the achievement. Expert 3 neither produces all possible knowledge states, nor does he attribute the learning material all to one skill.

Interpreting the first ("main") set of skill assignments, Expert 4 forms block of problems with equal difficulties and a few predictions about relations between pairs of problems. Expert 4 was the only one, who used the option of producing more than one set of skill assignments to the problems. Unfortunately, the resulting knowledge structure is not testable, because the prediction using the full sets of skill assignments of Expert 4 is almost the full Boolean Algebra of problems sets (only one relational information between 2 items remained). Therefore, we have used only the first set of skill assignments of Expert 4 for the subsequent data analysis.

The skill assignments of the four experts differ remarkably: For any two arbitrary problems, we observe a maximal agreement of at most two experts in terms of skills minimally necessary to solve these problems.

Table 3 shows some relational consequences of the skill assignment procedure in terms of the relation among items. In our particular context, there are three cases:

- $A_i = A_j$ if the skills minimally necessary to solve A_i are the same to solve A_j.
- $A_i < A_j$ if the set of skills minimally necessary to solve A_i is a proper subset of the set of skills minimally necessary to solve A_j.
- $(A_i, A_k) \to A_j$ if the union of skills assigned to A_i and A_k is a superset of the set of skills minimally necessary to solve A_j.

Every relation can be tested empirically by using the data introduced in the preceding section by dividing the number of subjects which account for the relation by the number of all subjects (Column 3 of Table 3). Whereas the prediction quality of expert 3 is satisfactory, the results produced by the other experts turn out to be disappointing.

Table 3: Restrictions within the knowledge structures given by the experts' skill assignment

	Restrictions	Evaluation
Expert 1	$A_1 = A_4 = A_9 = A_{10} = A_{15} = A_{17}$	0.333
	$A_5 = A_{11}$	0.708
	$A_7 = A_8$	0.917
	$A_3 = A_{14}$	0.875
Expert 2	$A_1 = A_2 = \ldots = A_5 = A_6$	0.292
	$A_7 = A_8$	0.917
	$A_9 = A_{10} = A_{11} = A_{12}$	0.250
	$A_{13} = A_{14}$	0.875
	$A_{15} = A_{16} = A_{17} = A_{18}$	0.500
Expert 3	$A_1 = A_9$	0.833
	$A_5 = A_{11}$	0.708
	$(A_5, A_{11}) \rightarrow A_2$	0.875
	$(A_5, A_{11}) \rightarrow A_{17}$	0.708
	$A_4 < A_{10}$	0.792
	$A_{14} < A_{13}$	0.875
	$A_8 < A_7$	1.000
	$(A_3, A_5) \rightarrow A_2$	0.917
	$(A_3, A_{17}) \rightarrow A_2$	0.958
	$(A_3, A_{11}) \rightarrow A_2$	0.917
Expert 4	$A_1 = A_4 = A_5 = A_{10} = A_{11}$	0.208
	$A_2 = A_3 = A_{12} = A_{13} = A_{15} = A_{16}$	0.292
	$(A_2, A_{17}) \rightarrow A_{18}, (A_3, A_{17}) \rightarrow A_{18}, (A_{12}, A_{17}) \rightarrow A_{18}, (A_{13}, A_{17}) \rightarrow A_{18}, (A_{15}, A_{17}) \rightarrow A_{18}, (A_{16}, A_{17}) \rightarrow A_{18}$	1.000 each

4 Summary of Conclusions

Knowledge assessment using the empirically determined knowledge structures offers interpretational variations: First, the empirical knowledge structure analysis highlights the internal (relational) structure of the knowledge of subjects, the relation between the problems, and finally the birelational structure of subjects and problems. Furthermore, differences between groups can be described by the empirical knowledge structure in more detail than by a summary statistic based on mean values of solved test items. This knowledge structure can be achieved by a few algorithmic steps, and an implementation of such a tool in a CALL system is therefore no substantial additional effort.

Asking experts to build knowledge structures based on a skill oriented knowledge description is a difficult task. Experts differ in their opinion, and the theoretical knowledge structure predicted by the experts often differs substantially from the empirical structure. The experts' and empirical data show that there are situational factors which mediate the 'pure knowledge' description.

One might argue that the results of the experts queries are of no interest, because either the chosen experts are not experts or the query procedure used is too demanding for a human processor. However, we firstly note that one subject reproduced most of the interesting relational information within the empirical knowledge structure; at least for this subject the results indicate that the procedure was not too demanding. Secondly, we think that it is quite impossible to achieve the results without expertise, simply by intuition. Nevertheless, the objection that the procedure is too demanding for the 'modal' expert has to be taken seriously. We have observed in earlier work that even in fairly structured contexts – like simple problem solving tasks in basic geometry – the problem analysis done by didactic experts is often incomplete. Therefore, we have to admit that the applicability of the problem analysis using our expert query procedure in a language achievement test is an interesting – but yet unsolved – problem.

5 Acknowledgement

Work on this paper was partially funded by the European Commission's RECALL project, Telematics Applications of Common Interest – Language Engineering, LE1-1615.

References

Düntsch, I. & Gediga, G. (1995). Skills and knowledge structures. *British Journal of. Mathematical and Statistical Psychology*, **48**, 9–27.

Düntsch, I. & Gediga, G. (1996). On query procedures to build knowledge structures. *J. Math. Psych.*, **40**, 160–168.

Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P. & Johannesen, J. (1990). Introduction to knowledge spaces: How to build, test and search them. *Psychological Review*, **97**, 202–234.

Fox, D. & Hamilton, S. (1996). Report on effectiveness study. RECALL (Telematics Project LE1 – 1615), deliverable 23, Commercial in confidence.

Koppen, M. & Doignon, J.-P. (1990). How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology*, **34**, 311–331.

Wille, R. (1982). Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival (Ed.), *Ordered sets*, vol. 83 of *NATO Advanced Studies Institute*, 445–470. Dordrecht: Reidel.