

An Intelligent Tutoring System for Learning Chinese with a Cognitive Model of the Learner

Michał Kosek & Pierre Lison

Language Technology Group, Department of Informatics
University of Oslo, Norway.

- How to represent *uncertain information* about the learner's vocabulary knowledge?
- How to use this information to *select* the most appropriate exercises?
- *Graphical models* provide a principled way of answering these questions.

Tutoring system

Please translate the following sentence into Chinese:
This woman listens to music while applying makeup

Don't panic! If you can't write the whole sentence, just write the words that you know. You'll get some tips how to improve it.

You can **click** on the **words** to look them up in a dictionary.

Many different translations are accepted as correct.

If you're not getting anywhere, despite trying, click Skip (I give up).

If the system doesn't accept your translation, but you know it's OK, click Skip (I'm sure my translation is correct).

Explanation of the feedback:
Green words are correct, but may still be in wrong order.
Grey parts may be correct or may be not, depending on how you will formulate the rest of the sentence.
Red parts need to be written in a different way.

Current sentence: 1 out of 25

Enable Chinese Input Method — click if you have no Chinese keyboard installed on your computer.

Dictionary · Google Translate · Help

Search by Chinese, Pinyin or English Definition:

Did you mean: *make up makebi* ?

结构 jié gòu structure / composition / makeup / architecture / CL:座.个
妆 zhuāng (of a woman) to adorn oneself / makeup / adornment / trousseau / stage makeup and costume
化妆 huà zhuāng cosmetic / makeup product
品 pǐn composition / make-up / ingredient / element / component / one's social status / CL:个
成分 chéng fèn to put on makeup
版面 bǎn miàn space of a whole page / layout (makeup) of a printed sheet
粉色 fēn sè pink / white / erotic / beautiful woman / powdered (with make-up) / composition / make-up / ingredient /

- Sequence of *online translation exercises* from English to Chinese
- Repository of 100 exercises, from beginner to advanced level
- Large number of accepted translations (>1 000 for some exercises)
- Nearest translation determined by calculating the *BLEU* score (standard metric employed for the evaluation of machine translation)
- *Interactive feedback* on frequent errors, missing or superfluous words, leads the learner towards one of the possible translations
- *Dictionary look-up* by clicking on underlined words

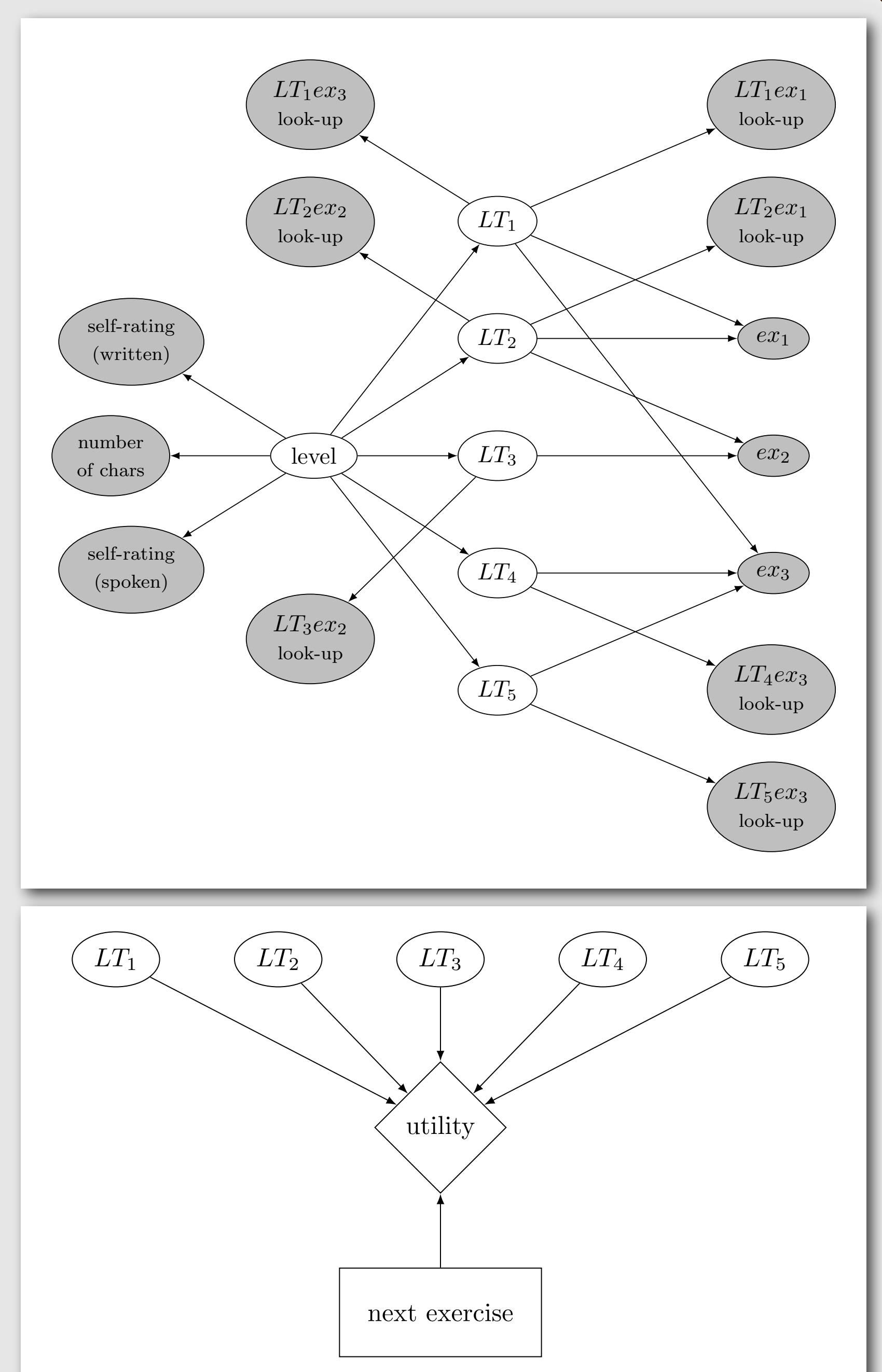
Evaluation

	User level	System (15 users)		Baseline (9 users)		p-value
		Mean	Std. dev.	Mean	Std. dev.	
Post-test results (number of correct answers, out of 18)	A	15.33	2.81	14.25	1.48	0.4963
	B1	16.75	1.09	17.75	0.43	0.215
	B2	18.00	0.00	18.00	0.00	n/a
	C	18.00	0.00	n/a	n/a	n/a
Users' subjective difficulty assessment ("way too easy" = 2, "a bit too easy" = 1, "OK" = 0, "a bit too hard" = 1, "way too hard" = 2)		0.67	0.79	0.56	0.50	0.6895
Users' subjective assessment of the number of items they have learnt ("none at all" = 0, "just a few" = 1, "some" = 2, "a lot" = 3)		1.53	0.88	0.89	0.57	0.04954
Drop-out rate (Fraction of users who did not finish the exercises)		0.38	0.49	0.31	0.46	0.6451

(Post-test results must be compared separately for every user level)

Cognitive modelling

- The learner model represented as a *Bayesian network*
- *Learning targets (LT)* are the words and constructions taught by the system
- Learning targets and the user level are *hidden variables*
- For each LT, the system stores the probability that the target is known by the user
- These probabilities are regularly updated based on available evidence:



- *Estimated number of known Chinese characters*
- *Self-rating of language level (oral and written)*
- *Exercise outcomes (success, failure, skip)*
- *Dictionary look-ups*
- In the decision network, the resulting utility depends both on the selected (next) exercise and the current vocabulary knowledge
- The *Zone of Proximal Development (ZPD)* is the set of tasks that the student cannot do alone, but can do given assistance from the system
- The utility is highest for exercises likely to lie in the student's ZPD (e.g. those that contain a few unknown words, but not too many)

- The system has been compared against a baseline that simply selects exercises at the user's declared level, without the cognitive model
- **Objective evaluation:** post-test on the user's vocabulary after the session
- **Subjective evaluation:** users' assessment of the difficulty of the exercises and the learning outcome
- No statistically significant difference between the system and the baseline for the post-test and the subjective difficulty assessment
- Subjective assessment of the learning outcome significantly better for the system than for the baseline ($p < 0.05$)
- Lack of significance may be due to too small sample sizes, experiments with larger groups of users left for future work