

UiO • Department of Informatics
University of Oslo

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

Michał Kosek

Master's Thesis Autumn 2014



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

Michał Kosek

August 15, 2014

Abstract

This thesis presents an intelligent tutoring system that enables students of Chinese to acquire active knowledge of words and grammatical constructions. The system relies on a Bayesian, linguistically motivated cognitive model that represents the estimated knowledge of the learner. This model is dynamically updated given observations about learner's behaviour and proficiency in the exercises. The model is then employed at run-time to select the exercises that are expected to maximise the learning outcome.

The system was implemented together with a set of 100 English-to-Chinese translation exercises. Each exercise is associated with a set of solutions. If the student's answer is not correct, the system finds a solution with the shortest distance to the input and gives interactive feedback using a combination of error-specific and generic rules that provide relevant cues towards the closest correct translation. The system is integrated with a bilingual English-Chinese dictionary, and the student may look up unknown words at any stage. Both dictionary look-ups and learner's proficiency in the exercises serve as evidence that enables the system to infer probabilistic information about the learner's actual vocabulary knowledge.

Compared with a baseline that randomly chooses exercises at the user's declared level, the proposed approach has shown a positive, statistically significant effect on the users' assessment of how much they have learnt. The results suggest that the cognitive system leads to improved learning outcomes. Experiments with larger groups of participants are required to detect potential differences in other effects.

Acknowledgements

In the first place I would like to express my gratitude to my supervisors: Pierre Lison, Ivo Spira and Jan Tore Lønning, who have offered me a great deal of advice from their areas of expertise, which had a significant impact on my thesis. Pierre Lison offered much support to my project proposal from the very beginning, agreed to supervise this thesis and took initiatives to overcome certain formal difficulties. He was very involved in the supervision, offered a lot of encouragement, advice and a wide range of feedback on many levels, from very detailed to very high-level. Ivo Spira offered many important and insightful comments related to Chinese and linguistics, which had a big influence on the thesis. Jan Tore Lønning gave me important insights into the methodology of conducting experiments and analysing their results.

The completion of this thesis would have been impossible without the help of my fellow second language speakers of Chinese. I am very grateful to those who used the tutoring system, making its evaluation possible. I want to first thank Eirik Tengesdal, Jo Flottorp Ness, Morten Aase Løver, Petter Mæhlum, Silje Andrea Berger Wahl and Victor Berrjod who have tested the system several times at various stages of its development and offered advice that helped improve later revisions and ensure that the experiment was conducted smoothly. I would like to thank everyone who took the time to try the system: Anna Lisa Ahlers, Bjørn Theisen, Federico Valenti, Gøran Olaussen, Hrafnkell Guðmundsson, Ingrid Winther Øyslebø, Isak Dahle, Kim Jarle Wroldsen, Lars Engen, Linn Christiansen, Luka Barl, Mads Melvold, Maja Husdal, Margrethe Knudsen, Michael B. Berthelsen, Oda Røste Odden, Paul Lobas, Rosa Partini, Sebastian Wojciechowski, Silje Christine Andersen, Tonje Kjellevold and others that I could not identify. The data obtained from the experiment and the feedback from the users play an important role in this thesis.

I am indebted to Wei Xuan 魏璇, Chang Xiaoyu 常小玉 and Zhang Lanyun 张兰云, who have spent many hours offering me details on differences between synonyms, various ways of expressing the same meaning, acceptability judgements and other native speaker insights that allowed me to create and revise most of the patterns and rules in the exercises. Without their help it would not be possible to make the exercises truly useful for language learners.

Finding enough second language Chinese speakers to take part in an experiment is not an easy task. I would like to thank Agnieszka Figiel, Aleksandra Piech, Henrik Fjellanger, Michał Belczyk, Michał B. Paradowski

and Mona Fonnes for helping me with that.

I am extremely grateful to John Gillooley, who has corrected my English in most parts of this thesis, and to Steve Pepper and Victor Berrjod, who have corrected the rest of it at short notice.

My thanks also go to the Text Laboratory at the University of Oslo for hosting the tutoring application, which facilitated its development and conducting the experiment.

Some of the ideas behind this thesis were conceived many years ago. I want to thank Piotr Woźniak, the creator of SuperMemo – an algorithm and a computer program that got me interested in the great potential of technology-enhanced tools for language learning. SuperMemo-based programs have been significantly contributing to my learning of Norwegian and Chinese since 2006; I would not be able to achieve my current level of Chinese without them.

Finally, I would like to give thanks to my girlfriend Martyna for the love and support during the difficult days of writing this thesis.

Contents

1	Introduction	1
1.1	Motivation	1
1.1.1	Student modelling	1
1.1.2	Design of a vocabulary tutor	2
1.1.3	Chinese as a foreign language	3
1.2	Outline	3
1.3	Theories, Approaches and Techniques	5
2	Background	7
2.1	Second Language Acquisition	7
2.1.1	Learning theories	7
2.1.2	Comprehensible input hypothesis	10
2.1.3	Noticing hypothesis and pushed output hypothesis	11
2.1.4	Feedback	12
2.1.5	Teaching approaches	13
2.2	Important Features of Chinese	14
2.2.1	Characters	15
2.2.2	Word segmentation ambiguity	15
2.2.3	Classifiers	16
2.2.4	Sentence structure	17
	Less function words and lack of inflection	17
	Basic word order	17
	Word order with co-verbs	18
	Word order with topicalisation	19
	Null arguments	20
2.3	Computer-Assisted Language Learning	20
2.3.1	Vocabulary tutors	21
2.4	Intelligent Tutoring Systems	22
2.4.1	Domain model	23
2.4.2	Student model	25
	Artificial intelligence-based methods	25
	Cognitive science-based methods	26
	Memory models	27
2.4.3	Tutor model	27
2.5	Summary	28

3	Approach	31
3.1	Rationale and General Design	31
3.1.1	Exercise types	31
3.1.2	Translation exercises	33
3.1.3	Feedback	34
3.1.4	Evidence	35
3.2	Directed Probabilistic Graphical Models	35
3.2.1	Representation of Bayesian networks	36
3.2.2	Inference in Bayesian networks	37
3.2.3	Noisy functional dependence	38
3.2.4	Utility theory and decision networks	39
3.3	Formalisation	41
3.3.1	Domain model: Constructions in sentences	41
	Constructions	41
	Multi-word constructions	42
	Learning targets	42
	Representation	43
3.3.2	Student model: Probabilistic information about learner's knowledge	44
3.3.3	Tutor model: Selecting the next exercise	46
3.4	Summary	47
4	Implementation	49
4.1	Overview of the Implementation	49
4.2	Natural Language Processing	50
4.2.1	Word segmentation	50
4.2.2	Sentence distance metrics	51
4.3	Character Recognition Test	52
4.4	Design and Implementation of the Client	53
4.4.1	User interface	53
4.4.2	Interactive feedback	54
4.4.3	Representation of the answer set	56
4.4.4	Calculating the sentence distance	56
4.5	Relational Representation of the Bayesian Network	57
5	Experiment	59
5.1	Testing and Data Collection	59
5.2	Parameter Estimation	60
5.2.1	User levels	60
5.2.2	Conditional probabilities	63
5.2.3	Utility values	65
5.3	User Evaluation	66
5.3.1	Experimental setup	66
5.3.2	Baseline	67
5.3.3	Results	67
5.4	Discussion	69
5.5	Summary	71

6	Conclusion	73
6.1	Summary of the Thesis	73
6.2	Future Work	74
6.2.1	Domain modelling	74
6.2.2	Memory modelling	75
6.2.3	Exercise encoding	76
6.2.4	Modelling user errors	76
6.2.5	Improved sentence selection	77
A	Exercises	79
B	Synonyms	97
	Bibliography	99

List of Figures

2.1	Typical relationship between models in intelligent tutoring systems	23
3.1	Example Bayesian network	38
3.2	Example decision network	40
3.3	Example domain model represented by a semantic network .	43
3.4	Example student model represented by a Bayesian network .	46
3.5	Example tutor model represented by a decision network . . .	47
4.1	The structure of CHINESE IN CONTEXT tutoring system . . .	49
4.2	Ongoing session with the program	55
5.1	Skip ratio, estimated number of known characters and self-assessment of written proficiency	61
5.2	Look-up ratio, estimated number of known characters and self-assessment of written proficiency	61
5.3	Look-up ratio, estimated number of known characters and self-assessment of spoken proficiency	62
5.4	Skip ratio, estimated number of known characters and self-assessment of spoken proficiency	62

Chapter 1

Introduction

Learning new languages is a time-consuming task. If we set very related languages and language savants aside, achieving even just an intermediate level in any language requires at least a few hundred hours, and in many cases the numbers are much higher.

Both the learning time and the ultimate achievement vary much more in second language learning than in first language acquisition. This suggests that the choice of learning strategy is important. There is no consensus about *the* proper strategy of learning foreign languages, and the diversity of approaches that have been proposed makes it very unlikely that there is one way of learning that is unconditionally better than others.

The new possibilities for creating language learning software tools are researched within the field of **computer-assisted language learning** (CALL). The advance in technology allows for creation of tools that can adapt to the learner to a much higher degree than before. How such user adaptations can be practically achieved is an area of ongoing research in CALL.

1.1 Motivation

This thesis presents the design, implementation and evaluation of an intelligent tutoring system (ITS) for learning Chinese with a cognitive model of the learner. The present section motivates why this kind of system is an interesting research topic.

1.1.1 Student modelling

The most important topic of the presented thesis is the **student model** (also called **learner model**). This is the part of a language tutoring system that stores information about an individual user, which can then be used to make the system adapt to the learner's needs. This thesis investigates a particular representation for this student model (Bayesian networks) and its empirical effects on learning outcomes.

When the system knows the state of the learner's current knowledge, it can provide tutoring at the right level and lead to more learning. This is a

major advantage of tutoring systems over a teacher in a classroom with many students. Compared with one-to-one human tutoring, the advantage of ITS is admittedly smaller. Still, when faced with much information, people often have problems making decisions. Even though tutoring remains an AI-complete problem (McCalla 2010), we can probably find many restricted areas of language teaching where an ITS with sufficient information about the learner can outperform an individual tutor.

It is not obvious how a student model should be structured. We can expect that a good student model should be **cognitive**: it should aim to describe what actually is in the learner’s mind, at some level of abstraction. However, there is no direct way to find out what the user knows, and how this information is structured. The system can only obtain indirect evidence about the user, based on the user’s behaviour. This means that the information that the system can obtain about the user’s current knowledge is uncertain. This is the challenge that this thesis attempts to address.

1.1.2 Design of a vocabulary tutor

Most vocabulary tutoring software that is presently in use does not fully exploit the possibilities that are opened by the advance of the technology. The words are often taught out of context, which is generally a bad strategy, which can work well only for specific nouns that have relatively stable referents across languages. If the words are taught in context, it is usually by means of traditional types of exercises, such as filling in gaps or putting words in a correct order. These exercises were designed with limitations of paper textbooks in mind. Technology overcomes some of these limitations, and therefore we should take a fresh look at what kind of exercises are most useful and possible to implement in language teaching software.

Some vocabulary tutors lack a student model. There are, however, a growing number of **spaced repetition systems** (SRS) that try to model user’s knowledge and find an optimal schedule of repetitions of exercises. The problem with these systems is their assumption that there is exactly one vocabulary item per exercise. Therefore, they are optimised towards learning words out of context. They can be used for learning whole sentences, but then a sentence is treated as a black box without any internal structure, and without any relation to other sentences. This is something that may lead to inappropriate selection of exercises. This problem is partially based on the fact that SRS are usually designed as tools to acquire any knowledge, without taking into account the specific needs of language learning.

This thesis presents a design of a vocabulary tutor with exercises that take advantage of language technology, and are motivated by the theories of learning and second language acquisition. The aim is to teach words in context, and model user’s knowledge in a way that takes into account the linguistic content of the exercise.

1.1.3 Chinese as a foreign language

The main topic of this thesis – the cognitive modelling of the learner in an intelligent tutoring system – is applicable to teaching any language. However, the secondary goal was to create a proof-of-concept application that can not only be used to evaluate the cognitive modelling, but will also actually be useful for language learners. Choosing Mandarin Chinese as the target language allowed for the opportunity to take advantage of features of Chinese that make natural language processing easier (e.g. lack of inflection), and make the program more interesting and useful, still being able to create it within the scope of a Master’s thesis.

The choice of Chinese requires some explanation. It is the language with the largest number of native speakers, but it is not nearly as popular as a second language: the number of foreign language learners of Chinese is far behind English, Spanish or French. The numbers are, however, increasing, and we can expect the importance of Chinese as a foreign language (CFL) to increase in the foreseeable future. More importantly, it is clear that Chinese, especially the written language, poses a great challenge to Western learners: Hayes (1987) found that while for Chinese natives the complexity of the characters did not influence their recognition, for second language learners of Chinese it did. Of course, the challenge of learning Chinese is much lower e.g. for Japanese. Therefore, the tutoring system presented in this thesis, is aimed at Western learners, who are most likely to speak English, which we assume to be the source language that the learners already can speak.

Apart from the complexity of the characters, Chinese words are very different from those of Indo-European languages: there are no cognates, and the number of loanwords is relatively small. This makes Chinese words much harder to remember for speakers of Western languages. For such learners, the risk of eventual failure in learning Chinese is much higher than in the case of learning a related language. Therefore, the question of choosing the right learning strategy becomes more important. This is where intelligent tutoring systems may help. As Ling (2007, p. 76) argues in her review of Chinese script acquisition studies: “Utilization of computer technology to increase learning effectiveness” is one of the main recommendations for future research in Chinese second language acquisition. This is not surprising, as technology has already provided large benefits to CFL learners. For example, the input methods on mobile phones that allow a user to look up unknown words and characters results in the saving of a huge amount of time, compared with looking them up in a paper dictionary. We can expect that intelligent tutoring systems that facilitate time-consuming tasks, and suggest useful, user-adapted learning strategies, can make learning Chinese significantly easier.

1.2 Outline

There are three questions that this thesis aims to investigate:

1. What can we learn from theories of learning and second language

acquisition for the development of language tutors?

2. How to structure the student model of a language tutor as a Bayesian network representing the learner’s vocabulary knowledge, and use this model to select the most appropriate exercises?
3. Does using this student model lead to measurable improvements in the learning outcomes of this tutoring system?

Below we provide an overview of the structure of the thesis, that indicates how these questions are answered. These questions are also addressed in two papers created as the result of this thesis: Kosek (in press) and Kosek & Lison (in press).

Chapter 2: Background

This chapter introduces concepts and provides information that is used in subsequent parts of the thesis. The chapter has four main sections. The first section provides important information related to learning theories, second language acquisition hypotheses, and teaching approaches. Among others, it introduces the zone of proximal development (ZPD), the pushed output hypothesis, dynamic corrective feedback and the lexical approach. The second section introduces important features of Chinese language that might not be obvious for a speaker of English. It briefly describes how characters are combined into words, what classifiers are and what a typical Chinese sentence structure is. It also details which features may cause problems for processing Chinese by machine. In the third section we introduce the field of computer-assisted language learning, and more specifically vocabulary tutors. The fourth section describes the structure of intelligent tutoring systems (ITS), which contains three important models: the domain model, the student model and the tutor model.

Chapter 3: Approach

The chapter presents our approach to designing a vocabulary tutor and tries to address the first two questions posed above. The first section uses the information from the background chapter to motivate a general design of a vocabulary tutor: the types of the exercises, the way feedback should be provided and what kind of evidence can be made available to the student model. This is where we present an answer to the first question. The next section introduces two probabilistic graphical models (Koller & Friedman 2009): Bayesian networks and decisions networks, which are tools that will be used in connection with the core subject of this thesis, indicated in the second question: student modelling. The application of a Bayesian network in the design of the student model is presented in the third section, along with the design of the two other ITS models.

Chapter 4: Implementation

This chapter presents an actual implementation of our approach. It provides information about natural language processing techniques that were used in the implementation, such as word segmentation and sentence distance metrics. The chapter shows how an actual interaction with the system looks, and some details about implementation of some of the components.

Chapter 5: Experiment

The chapter provides information on how the system was tested and how its parameters were estimated. Most importantly, it presents an experiment that was conducted, and discusses its results. The system was found to be performing as well as a baseline that did not use a student model, and answers to one of users' self-evaluation question showed a positive, statistically significant difference between the system and the baseline that suggests that the system with a cognitive student model leads to more effective learning.

Chapter 6: Conclusion

This chapter provides a summary of the thesis, and presents suggestions for future research directions.

1.3 Theories, Approaches and Techniques

Creating a language tutoring system certainly requires using theories, approaches and techniques based on research in both linguistics and computer science. This section provides a short overview of how theories are used in computer-assisted language learning, which lies at the intersection of these disciplines.

Levy & Stockwell (2006) noted differences between research in CALL and in general second language acquisition (SLA): the field of SLA seems more unified, with more focus on acquisition of grammar and on general language acquisition theories, while in CALL there are separate lines of research for teaching pronunciation, grammar, vocabulary and discourse, applying theories that are specific to their particular field. They also point out that learning theories are usually created with researchers in mind, but in CALL there are two other groups that use them: CALL designers and language teachers. While researchers tend to use only one theory, in CALL design and applications of CALL for language teaching it is common to mix multiple theoretical perspectives.

This thesis is concerned with a design of a tutoring system that can be used in practice for language learning, whose way of teaching is consistent with what is known about second language acquisition, and that uses artificial intelligence (AI) and natural language processing (NLP) techniques to implement it. These goals clearly require drawing on multiple theories. The most important theories, hypotheses and approaches that

informed the design of the system and provided a framework to analyse it include: Vygotsky's sociocultural learning theory (subsection 2.1.1), Swain's pushed output hypothesis (subsection 2.1.3) and Lewis's lexical approach (subsection 2.1.5). The most important AI tools used in the system are probabilistic graphical models, namely Bayesian networks (subsection 3.2.1) and decision networks (subsection 3.2.4). The implementation uses NLP techniques, that involve word segmentation (subsection 4.2.1), computing distance between natural language sentences (subsection 4.2.2) and a formal grammar used for encoding the set of possible translations (subsection 4.4.3). The design of the module that provides feedback to the user is informed by the structure of dialogue systems (subsection 4.4.2).

Chapter 2

Background

This chapter has four main sections. First, we will look at theories and hypotheses related to learning and second language acquisition, which will give us important insights for constructing a tutoring system. The second section has two purposes. On the one hand, it will discuss features of Chinese that may cause problems for English speakers, and are therefore important to be taught by a language tutor. On the other hand, it will point out what makes natural language processing of Chinese texts different from processing English. Therefore, we will look at some important features of the Chinese language, stressing those that are different from English. In the third section we will introduce the field of computer-assisted language learning and discuss issues related to vocabulary tutors in more detail. Finally, we will take a look at issues related to designing intelligent tutoring systems and their components.

2.1 Second Language Acquisition

In this section we will discuss some important concepts of learning theories, giving much weight to the zone of proximal development. Then we will concentrate on hypotheses related to second language acquisition: comprehensible input hypothesis, noticing hypothesis and pushed output hypothesis. Then we will discuss issues related to giving feedback during second language learning. We will conclude with a quick look at language teaching approaches, and argue why the lexical approach can be used to motivate parts of this thesis.

2.1.1 Learning theories

This thesis is concerned with tutoring systems. On the most general level, the aim of a tutoring system is to help the user to learn new knowledge and skills. Before we begin discussing specific issues related to learning languages, let us look at learning in general. In educational science, **learning theories** are frameworks that try to provide a model of how people learn.

Three important paradigms that shaped learning theories in the 20th

century are **behaviourism**, **cognitivism** and **constructivism**. According to behaviourists, teaching amounts to eliciting appropriate responses to given stimuli, without any concern to learner's mental states. This approach was seen as too simplistic by cognitivists, who acknowledged the important role of learner's mental processes in learning. This paradigm emerged in 1950s, at the time when the field of artificial intelligence (AI) started taking shape, and adopted the view that learning, just like other mental processes, is essentially information processing. Consequently, teaching was seen as structuring information in a way that is easiest to process by the mind, taking into account various aspects of cognition, such as processing of sensory input, short- and long-term memory. The learner was seen as a passive receiver of properly structured information (Leonard 2002).

The third paradigm, constructivism, which entered the field of educational technology in the 1980s, postulates that the learner has a much more active role in the process. Learning works best, according to constructivists, when it is based on solving problems that build on learner's prior knowledge and experience, which is different for every individual. Also ways of integrating new knowledge obtained by problem solving vary between learners. Therefore, there is no objective knowledge that is transmitted from the teacher to the learner, as believed by cognitivists – the knowledge is constructed by every learner during active involvement with problem solving (Selwyn 2011).

The aforementioned paradigms are present in most modern learning theories to various degrees. In this thesis we will employ concepts taken from a constructivist **sociocultural learning theory**, which acknowledges that learning has both cognitive and social aspects. This theory is based on the work of a Soviet psychologist Lev Semyonovich Vygotsky, and has been gaining popularity since the publishing of an influential English translation of his works (Vygotsky 1978). One of the key concepts in Vygotsky's theory is the **zone of proximal development**. It is defined as

the distance between the level of current development, determined by problems that are solved independently [by the child] and the level of possible development, determined by problems that are solved by the child under adult guidance and in collaboration with more capable peers (Vygotsky 1935, p. 42).

In other words, the tasks in the ZPD are the ones that the learners cannot do by themselves, but can do if they are provided necessary help. The non-trivial insight in Vygotsky's theory is that it is the ZPD that is relevant to learning, not the actual developmental level. In other words, two children on the same developmental level may have very different zones of proximal development, and therefore require very different teaching strategies. The actual developmental level is therefore not interesting in of itself, but only as a measure of the lower bound of the ZPD.

The process of assisting learners has been called **scaffolding** (Wood, Bruner & Ross 1976). The initial focus of the learning theories was on children, hence Vygotsky's reference to the adult guidance in the definition of ZPD. However, we will see below that they have a wider application.

It should be noted that these ideas are not specific to Vygotsky's sociocultural theory. There is a parallel between the ZPD on the one hand, and **optimal mismatch** on the other, the latter being based on another influential constructivist learning theory developed by Jean Piaget (Pass 2004).

Let us now look at the role of the teacher in identifying ZPD and providing scaffolding. Tharp, Gallimore, et al. (1988) list four stages of progression through and beyond ZPD:

- In Stage I, the learners are **other-regulated**: they cannot do the task by themselves, and rely on support from the tutor. The tutor's assistance is responsive – the learners are gradually given more freedom in performing the task as their skills progress.
- Stage II begins when the learners become **self-regulated** and can perform the task without external help. This does not mean, however, that the skill is fully automatised, but only that the learners can notice and correct their own mistakes without assistance in a process that may involve self-directed speech.
- In Stage III, the self-regulation is not needed, and the learners have internalised their skills. At this stage, the learner progresses out of ZPD. The skill becomes “fossilised” – the performance is now harder to change than before.
- As new situations arise, Stage IV may, and frequently does occur, in which the skill becomes de-automatised and adapted to a new context, in which the already internalised performance is not good enough.

Let us now look at ways of helping learners in Stage I. There are 6 “scaffolding functions” listed by Wood, Bruner & Ross (1976, p. 98):

- “Recruitment”. The tutor should get learners interested and motivated to do the task.
- “Reduction in degrees of freedom”. The number of choices that the learner faces should not be too overwhelming, which may involve splitting the task into small, manageable steps.
- “Direction maintenance”. This involves making sure that the learner proceeds towards completing the task, and may include giving encouragement, offering help and directions.
- “Marking critical features”. If the learner produces something that is not correct and does not lead toward solving the problem, the tutor should let the learner know, and possibly offer hints as to how it should be improved.
- “Frustration control”. This function is summarised by Wood, Bruner & Ross (1976, p. 98) with the maxim: “Problem solving should be less dangerous or stressful with a tutor than without”.

- “Demonstration”. If the learner is not able to finish the task, despite the help, the tutor should demonstrate how to solve it, in a way that can later be imitated by the learner in subsequent tasks.

We can see that these functions are quite general, and can be used in pedagogical approaches that apply to both children and adults. They can also serve as guidelines for developing tutoring systems.

Vygotsky stressed the fact that virtually all human interactions with their environment are **mediated** by tools. Tools include not only physical items, such as ploughs, hammers, or computers, but also symbolic ones, such as language. Language is indeed an important tool in most instances of teaching and learning. Laurillard & Marullo (1993) point out, however, that Vygotsky was concerned with the acquisition of one’s first language and its role in learning, and that the acquisition of foreign languages is different in terms of their role and the ways in which they are learnt. Therefore, we will now turn to theories that are specifically concerned with second language acquisition.

2.1.2 Comprehensible input hypothesis

There are several hypotheses relating to how people acquire foreign languages. Krashen (1985) posited **comprehensible input hypothesis**, which says that people learn by analysing input they receive, as long as they can comprehend it. A logical consequence of this hypothesis is that the most beneficial input is one that is slightly above student’s current level. Learning can only occur if there are some previously unknown elements in the input. There cannot be too many of them, however, because that would make the input incomprehensible. Such most beneficial input is often called $i+1$ -level input, where i is current level in some scale.

The notion of $i+1$ is often associated with Vygotsky’s concept of zone of proximal development (cf. subsection 2.1.1). However, Lantolf & Thorne (2007) point out several important differences. The $i+1$ level is shared across learners at a given development stage i , and is not predictable in advance. The content of ZPD is specific to a learner in a particular situation, and it is something that can be determined empirically during teaching. The empirical focus makes ZPD more directly applicable in pedagogy. It requires, however, a more nuanced approach, as it stresses the fact that two learners who are at the same level (according to some scale) do not necessarily have the same ZPD, and therefore may require different treatment.

Regardless of controversies surrounding how much human language faculty is innate, there are obviously many aspects of language that need to be learnt, and learning must ultimately come from input. Input may be in the form of implicit or explicit knowledge.

For first language speakers it is quite certain that the vast majority of words is learned unconsciously. For second language speakers it is much harder to assess, given the diversity of ways people acquire their second language. There is for example substantial amount of empirical evidence that supports the hypothesis that explicit learning techniques, such as

mnemonics, increase the recall of recently learned words (Koda 2005, p. 55). Conversely, Shen (2007) found out that advanced learners of Chinese acquire vocabulary through independent reading and that the vocabulary gain rate was negatively correlated with the number of characters that were unknown to the readers. In other words, if almost all characters in a text are known, it is much easier to learn new words from context.

Moreover, languages have many patterns without clear rules that can be stated explicitly, yet many learners manage to acquire such patterns. In such situations, the implicit knowledge in the input is the only possible source of learning. We can conclude that implicit knowledge always plays some role in the acquisition process.

If input contains only the patterns and constructions that are already mastered by the learner, there is no new information that could lead to increasing one's level. On the other hand, it is easy to observe that input with too many unknown constructions does not lead to any learning. If there is an optimal level of input that maximises learning in a given situation, it is certainly somewhere between the two, as predicted by the hypothesis.

2.1.3 Noticing hypothesis and pushed output hypothesis

There are, however, situations that are hard to account for with the comprehensible input hypothesis. Schmidt (1983) and Swain (1985) have shown learners who stopped learning at a quite low level and did not improve despite continuous access to comprehensible input at a higher level. This led Schmidt (1995) to form the **noticing hypothesis**, which assumes that learning a particular language feature must be preceded by some kind of conscious noticing of that feature in the input. Swain (1985) came up with the **pushed output hypothesis**: learning can be caused by pushing a learner to produce output that he did not produce before, and providing feedback afterwards.

The noticing hypothesis restricts the applicability of the input hypothesis. A construction or pattern can be learned with comprehensible input only in situations when the learner is able to notice this construction/pattern. Differences between one's L1 and L2 often cause some elements of L2 to remain unnoticed. Input carries only implicit knowledge, which is not enough – learning must therefore involve explicit knowledge. Explicit description of specific constructions may help to notice them in the input and, subsequently, learn them. This model, although reasonable and able to account for more cases than the input hypothesis, has an important disadvantage: it is based on the concept of noticing, which is in turn based on the concept of conscious attention. It is hard to test the noticing hypothesis scientifically or use it pedagogically, because it is difficult to examine directly whether something was consciously noticed. In this regard the pushed output hypothesis, which can be seen as a consequence of the more general noticing hypothesis, is more practical.

The following example will illustrate this difference. Let us say a learner of Chinese does not know the word 棕色 *zōngsè* “brown”, and somebody uses this word in an utterance that is understandable and it is clear from

the context that 棕色 *zōngsè* refers to that colour. According to the comprehensible input hypothesis, exposure to such sentences is a necessary and sufficient condition for internalising the word for “brown”. The noticing hypothesis paints a more nuanced picture: it might be the case that the learner understood the meaning of the sentence, but did not notice the form of the word 棕色 *zōngsè*. This can be caused by a number of factors, determined by e.g. learner’s native language and salience of the word in the utterance. If unfavourable conditions persist in other similar sentences containing the word 棕色 *zōngsè*, the word will not be learnt, despite the access to comprehensible input. Therefore information about what sentences were understood is not enough to conclude what words are likely to have been learnt. However, if a learner needs to produce the word in a sentence, they will succeed only if the word has been noticed before. If it has not been noticed and it is an important part of the sentence, they may use various strategies to find the word, e.g. look it up in a dictionary or, in the case of the word for “brown”, point to a brown object and ask the interlocutor what that colour is called. In any case, the word 棕色 *zōngsè* will become very salient in that situation, and will be likely to be noticed. We can therefore understand the pushed output hypothesis as describing externally controllable scenarios in which new words are likely to be noticed and learnt.

2.1.4 Feedback

An important aspect of the pushed output hypothesis is feedback. We can view it as a particular form of scaffolding, which was discussed in subsection 2.1.1.

Depending on whether unacceptability or acceptability is being signalled, it can be negative or positive feedback. Negative feedback may vary in the degree of explicitness: explicit corrections usually focus on form and may provide much information (a typical example might be a language teacher saying “*He is doctor*” is wrong, you should use an article before a noun: “*He is a doctor*”), while implicit feedback, e.g. clarification requests, focus on meaning, e.g. *sorry?* signals that the utterance has not been understood (Ortega 2013), but usually does not give information how it should be changed. Positive feedback may also have various degrees of explicitness, but the difference has less practical consequences, as it does not need to signal anything more than the acceptability of the utterance in terms of meaning and/or form.

Let us look closer at issues surrounding written corrective feedback. Its effectiveness is subject to controversy. A recent example is a meta-analysis performed by Truscott (2007), who obtained results that suggest that traditional forms of written corrective feedback (WCF) do not have any positive effect on students’ ability to write accurately. This does not mean, however, that feedback cannot have benefits. Evans et al. (2010) suggested that the right question to ask is not whether traditional written corrective feedback is effective, but rather what methods one can use to improve students’ writing accuracy. Hartshorn et al. (2010) tested the empirical effects of a novel form of WCF, that they called **dynamic written**

corrective feedback and found it to be effective. This form of feedback has four distinguishing properties:

- **Meaningful.** The learner must know what kind of error they made, and how it can be corrected. Meaningful feedback may not give the complete correct answer, but gives enough information so that the learner can correct the error herself.
- **Timely and constant.** The feedback should be provided as soon as possible, and its form and scope should not change, so as not to confuse the learner.
- **Manageable.** The feedback cannot be overwhelming. It needs to be provided in manageable chunks, and information overload should be avoided.

Some of these features are related to the scaffolding functions explained in subsection 2.1.1: manageability has to do with reducing degrees of freedom, and meaningfulness – with proper marking of critical features.

The above findings are consistent with the meta-analysis by Lyster & Saito (2010), who found that in oral contexts the most effective is **prompt feedback**, which does not provide an alternative formulation of the incorrect utterance, but provides cues that let the learner self-repair. We can see that kind of feedback is also likely to be effective in written contexts with high degree of interactivity.

2.1.5 Teaching approaches

Let us end this overview of important concepts in second language acquisition with a look at modern teaching approaches. Most current mainstream approaches to teaching foreign languages are based on a **communicative approach** that shifted emphasis from teaching language systems, such as grammar and vocabulary, to teaching how to communicate in the language (Thornbury 2006), with the assumption that the language systems will be eventually acquired implicitly. We have, however, seen in the previous subsections that there are some aspects of language that are hard to learn when focusing solely on meaning, and not on form.

Laurillard & Marullo (1993), who looked at second language learning from the perspective of Vygotsky’s sociocultural theory (cf. subsection 2.1.1), concluded that this theory does not support teaching with a sole focus on meaning. Ma & Kelly (2006) discuss the distinction between implicit and explicit learning from the perspective of vocabulary acquisition, and conclude that evidence is in favour of using both to some degree. Many proponents of communicative teaching recognise this and admit that systematic teaching with focus on form is sometimes needed (Thornbury 2006).

An important alternative approach, the **lexical approach**, can give us insight as to how such systematic teaching can be done. Unlike traditional approaches that emphasised grammar, the lexical approach focuses on vocabulary, claiming that “language consists of grammaticalised lexis, not

lexicalised grammar” (Lewis 1993, p. vi). The focus is not only on words, but also **chunks**, which can be any sequences of words that frequently occur together: collocations and formulaic language. Learning chunks that are frequent in native speakers’ speech is seen as a way to help the learners gain fluency and express themselves more idiomatically. The question of whether something is grammatical becomes therefore less important, and the focus is on whether it is idiomatic (i.e. whether native speakers actually speak that way).

The key idea of this approach can be summarised as follows: while traditional approaches see speaking as filling slots in some predefined sentence patterns, the lexical approach sees it as combining a set of predefined chunks into an utterance (Thornbury 2006). Thornbury (2013) notes also that proficiency in receptive skills requires large vocabulary, citing Laufer (1997) who found insufficient lexical skills as the greatest obstacle to reading.

We can see that in the lexical approach the issue of teaching vocabulary and larger multi-word items is given high importance, which can motivate greater research focus on vocabulary tutors. In the following three sections we will approach the issue of teaching words and phrases. First we will provide some basic information about the language we aim to teach, Chinese (section 2.2). Then we make an overview of the field of computer-assisted language learning (section 2.3), also covering issues that are specific to vocabulary tutors. In the final section of this chapter we will describe intelligent tutoring systems (section 2.4).

2.2 Important Features of Chinese

Chinese is often seen as a difficult language, and this opinion is not without its merits. This is a frequent opinion among Western learners, which is the group that we focus on in this thesis. Learning Chinese characters, for example, is certainly time-consuming when compared with most (if not all) other writing systems that are currently in use. Other reasons for the perception of Chinese as a difficult language is its linguistic distance from English and other Western languages.

Presenting all the important points of Chinese grammar is beyond the scope of this thesis, therefore this section presents only some of the aspects of Chinese that may be unknown to those who are familiar with English. It presents only those features of Chinese that are relevant for the description of the tutoring system in the next chapter. The features can be relevant in two different ways: it can be either a source of problems for the learners, and therefore something that should be taught by such a system, or it is something that may potentially cause problems with language processing by a machine.

Some of the differences between English and Chinese may make computational processing of Chinese more difficult, even though they are relatively easy for learners. For example, learners may often intuitively understand topicalisation in Chinese, even if it works in a different way

than in their native languages; however, the flexible word order caused by topicalisation definitely makes parsing Chinese sentences harder.

2.2.1 Characters

The most noticeable difference between Chinese and English is the use of characters instead of letters. We should note that while there are many frequent one-character words, most words in Chinese contain two characters, and in the overwhelming majority of cases individual characters represent morphemes.

Chinese native speakers know a few thousand characters: in the People's Republic of China, 2000 is considered a lower threshold for considering an urban resident literate, and 6500, according to the authorities, "satisfies needs related to publishing news, printing and editing".¹ We can therefore assume that a highly educated native speaker of Chinese who knows about 6 thousands characters can be fully proficient in all kinds of modern written language. Advanced second language learners of Chinese are expected to know about 2600-2800 characters, as this is the number of different characters in word lists used for preparation to an official Chinese proficiency exam (HSK, 汉语水平考试 *Hànyǔ Shuǐpíng Kǎoshì* "Chinese Proficiency Test").

For a language learner, this amount of characters poses a serious problem. One cannot learn the whole writing system before starting to learn vocabulary, which is a typical strategy for most other languages. Learning Chinese vocabulary is therefore harder, because new words often contain unknown characters. Both the complex structure of many characters and their sheer number cause problems for learners, such as confusing one character for another, and frequent forgetting. Hayden (2005) found that even advanced learners experience significantly higher cognitive load than native speakers during reading Chinese. All those factors show that an efficient vocabulary tutor that makes best use of learner's cognitive abilities, such as attention and memory, is a very important tool for learning Chinese.

From the computational perspective, the use of characters does not make processing any harder. There is, however, a related issue that does cause problems: word boundaries are not marked in Chinese in any way. In the following subsection the problem is discussed in more detail. Modern Chinese fortunately uses punctuation marks, which allow the easier separation of larger units, such as clauses and sentences.

2.2.2 Word segmentation ambiguity

Wong, Li, et al. (2009) cites the lack of word boundaries in Chinese as the main reason behind the difficulties of Chinese language processing in comparison with Western languages. According to Zong (2007), there are three main reasons why Chinese word segmentation is difficult: the ambiguity of the notion of word in Chinese, the need to recognise unseen words and ambiguity of the segmentation.

¹http://www.gov.cn/zwgc/2013-08/19/content_2469793.htm

The problem with defining what the word is in Chinese has two reasons. Before the 20th century the speakers of Chinese did not seem to have a concept of word at all, so word boundaries have never been used in writing. More importantly, Chinese morphology mirrors syntax much more than in other languages, so linguists often cannot use the cues that have been used to find boundaries in other languages that do not separate words in writing, or do not have a writing system at all. This is also related to the problem with recognising unseen words: new terms and acronyms can be created easily by compounding existing morphemes and there are not many restrictions on how they can combine. Similar problems are posed by proper nouns (names of people, places and organisations) that are not restricted either, and in general do not follow any pattern (such as capitalisation in the Western languages) that can allow to distinguish them from other words.

The ambiguity of the segmentation can be seen in the following example (Wong, Li, et al. 2009, p. 2). It shows two possible interpretations of the phrase 香港人口多: “The population of Hong Kong is large” or “Hong Kong people are talkative”.

香	港	人	口	多
<i>Xiāng</i>	<i>gǎng</i>	<i>rén</i>	<i>kǒu</i>	<i>duō</i>
Hong	Kong	population	many	
Hong	Kong	people	talkative	

In principle, one needs to understand the context of the utterance to resolve such ambiguities. However, even without the context, the morphological analyser may have the information that 人口 *rénkǒu* “population” is a frequent word, while 口多 *kǒuduō* “talkative” is a slang term restricted to Hong Kong. Therefore, the prior probability of the first interpretation is larger. Then, the statistical patterns in the context of the utterance may give an indication as to the correct interpretation, even without involving semantics – just like in the case of part-of-speech tagging.

2.2.3 Classifiers

Chinese also has **classifiers**, which form an important grammatical category that does not exist in English. They are often called **measure words**, although strictly speaking these are two different concepts. Measure words apply to uncountable nouns, and appear in English. For example, in the phrase *three loaves of bread*, *loaves* is a measure word. Classifiers are analogous words that apply to countable nouns. For example, while in English we would say *three people*, in Chinese we cannot simply combine the word 三 *sān* “three” with the word 人 *rén* “person”, we need to have a classifier 个 *ge* in between:

三	个	人
<i>sān</i>	<i>ge</i>	<i>rén</i>
three	<i>classifier</i>	people

个 *ge* is a general classifier that can be used with most nouns. Some nouns, however, do not accept 个 *ge*, and there are also many nouns that can be

used both with 个 *ge* and with some other, more specific, classifier. It should be noted that the choice of a classifier does not depend on a formal property of the noun, e.g. in the way gender determines articles in German. It may depend on the meaning of the noun (for example, 座 *zuò* is used with buildings, 辆 *liàng* is used with wheeled vehicles and 所 *suǒ* is used with institutions), but also may depend on the wider context (for example, in the context of counting number of people in a family, 口 *kǒu* is used instead of 个 *ge* in front of 人 *rén* “people”).

2.2.4 Sentence structure

Consider a task of recognising possible translations of a particular English sentence. Zong (2007) lists three main differences between Chinese and English grammar that we need to take into account:

- Chinese uses less function words and has no inflection.
- Word order in Chinese phrases is less regular than in English.
- Subject and object are not obligatory.

Let us look closer at these differences.

Less function words and lack of inflection

In Chinese, nouns are not preceded by articles, plural endings are used only for humans and are not obligatory, and verbs have only one form and require a preposition less often:

狼	来	了
<i>láng</i>	<i>lái</i>	<i>le</i>
wolf	come	PFV

The above sentence may have many different translations into English, depending on the context: “A wolf has come”, “The wolf has come”, “Wolves have come”, “The wolves have come”, “A wolf came”, “The wolf came”, etc. We can note the lack of inflection of verbs and nouns, and lack of obligatory definiteness markers, such as English articles *a* and *the*. 了 *le* is one of very few grammatical particles in Chinese. Its function may vary, depending on its place in the sentence. In this and all the following examples, it has a role of a *perfective aspect marker* (glossed as PFV) – it is used when referring to completed actions, and it is usually obligatory in such cases. In this function it is usually placed directly after the verb. When placed at the end of a sentence, it can also have this function, but usually fulfils other purposes, which will not be discussed here.

Basic word order

Basic word order in Chinese is: *who, when, where, what*, as opposed to English: *who, what, where, when*, as seen in the following example (Herzberg & Herzberg 2010, p. 19):

我们	昨天	下午	一	点钟	在
<i>wǒmen</i>	<i>zuótiān</i>	<i>xiàwǔ</i>	<i>yì</i>	<i>diǎnzhōng</i>	<i>zài</i>
we	yesterday	afternoon	one	hour	at
麦当劳	吃	了	午饭		
<i>Màidāngláo</i>	<i>chī</i>	<i>le</i>	<i>wǔfàn</i>		
McDonald's	eat	PFV	lunch		

“We ate lunch at McDonald’s at 1:00 yesterday afternoon”

As we can see, both English and Chinese are SVO languages, but the word order is not identical.

Word order with co-verbs

The word order in Chinese is more flexible than in English. One issue with word order is related to **co-verbs**. They are verbs that appear in a sentence in addition to the main verb, and have the role that is equivalent to the function of prepositions in English sentences. For example, in the following sentence, 道歉 *dàoqiàn* “apologise” is the main verb, and 给 *gěi* “give” is a co-verb that in this context can be translated as the preposition *to*:

他	给	我	道歉	了
<i>tā</i>	<i>gěi</i>	<i>wǒ</i>	<i>dàoqiàn</i>	<i>le</i>
he	give	I	apologise	PFV

“He has apologised to me”

Co-verbs usually precede the main verb, and cannot be moved, as is the case in the sentence above. For some main verbs, however, co-verbs can appear before the verb, between the verb and the object, or after the object:

我	给	他	打	电话
<i>wǒ</i>	<i>gěi</i>	<i>tā</i>	<i>dǎ</i>	<i>diànhuà</i>
I	give	he	call	telephone
我	打	给	他	电话
<i>wǒ</i>	<i>dǎ</i>	<i>gěi</i>	<i>tā</i>	<i>diànhuà</i>
I	call	give	he	telephone
我	打	电话	给	他
<i>wǒ</i>	<i>dǎ</i>	<i>diànhuà</i>	<i>gěi</i>	<i>tā</i>
I	call	telephone	give	he

“I will call him”

The same meaning can also be expressed without a co-verb, but with the indirect and the direct object following the verb:

我	打	他	电话
<i>wǒ</i>	<i>dǎ</i>	<i>tā</i>	<i>diànhuà</i>
I	call	he	telephone

“I will call him”

There are also other possible small changes to the sentence that do not change its meaning, enabled by the fact that nominalisation and

verbalisation is more frequent than in English, and does not require any change of the word form. In the following example the verbalisation is translatable into English:

我 给 他 电话
wǒ gěi tā diànhuà
 I to he telephone
 "I will telephone him"

However, this particular phrase is used less often by native speakers, as evidenced by a search in a Chinese search engine Baidu,² and is perceived as less correct than the previous ones.

Word order with topicalisation

Probably the most common mechanism that makes the word order more flexible in Chinese is **topicalisation**. Clause elements can be topicalised (moved to the beginning) in many more cases than in English. Let us look at a few examples, based on Xiong (2006), showing sentences both in default and topicalised word order.

Topicalisation of subjects and objects is possible both in English and Chinese:

我 不 相信 这些 话
wǒ bù xiāngxìn zhèxiē huà
 I not believe these words
 "I do not believe these words"

这些 话 我 不 相信
zhèxiē huà wǒ bù xiāngxìn
 these words I not believe
 "(As for) these words, I do not believe (them)"

In Chinese subordinate clauses can also be topicalised, in a way that is hard to translate into English:

我 不 相信 他 会 说 这些 话
wǒ bù xiāngxìn tā huì shuō zhèxiē huà
 I not believe he can say these words
 "I do not believe that he can say these words"

他 会 说 这些 话 我 不 相信
tā huì shuō zhèxiē huà wǒ bù xiāngxìn
 he can say these words I not believe
 "?That he can say these words, I do not believe"

In Chinese it is also possible to have more than one topic. In the following example, the second sentence has two topics: 昨天 *zuótiān* "yesterday" and 李先生 *Lǐ xiānshēng* "Mr. Li".

²<http://www.baidu.com>; "给他打电话": 12,900,000 results, "给他电话": 2,150,000 results [Retrieved 2014-07-26]

我 昨天 看见 了 李 先生
wǒ zuótiān kànjiàn le Lǐ xiānshēng
 I yesterday see PFV Li mister
 “I saw Mr. Li yesterday”

昨天 李 先生 我 看见 了
zuótiān Lǐ xiānshēng wǒ kànjiàn le
 yesterday Li mister I see PFV
 “?Yesterday, Mr. Li, I saw (him)”

Null arguments

In English, the presence of subject is usually obligatory. In the case of transitive verbs, object must be expressed as well. Chinese, on the other hand, is a **null-argument language**, which means that expressing subject and/or object is often not required. For example, as an answer to the question “Do you like her?”, one may say:

我 很 喜欢 她
wǒ hěn xǐhuan tā
 I very like she
 “I like her very much”

However, it is just as correct (and very frequent, due to brevity) to drop both the subject and the object:

很 喜欢
hěn xǐhuan
 very like
 “(I) like (her) very much”

The smaller number of function words and lack of inflection makes the task of processing Chinese easier than processing English, as there are less formal features to keep track of. On the other hand, the lack of inflection, combined with the relatively flexible word order and subject/object dropping, makes more sentences potentially valid, and makes it harder to analyse what functions words have in a sentence.

2.3 Computer-Assisted Language Learning

Computer-assisted language learning (CALL), also called **technology-enhanced language learning** (TELL), is a field concerned with any use of computers that leads to language learning, including e.g. language tutors, computer games and instant messaging programs.

An important distinction within CALL is that of the difference between CALL tutors and CALL tools. Tutors are standalone programs which directly aim at language teaching, while tools may be more general purpose. Levy & Stockwell (2006) describe tools as “enabling” devices, and give two groups of tools as examples: those related to computer-mediated communication (CMC), which enable better communication with other

learners and native speakers of the language, and those that facilitate access to language corpora and dictionaries of various sorts. One of the advantages of tools over tutors is that they can be easily adapted to various learning styles and promote the independence of the learner. Tutors, on the other hand, have some pre-defined syllabus. In section 2.3.1 we shall discuss the issue from the perspective of vocabulary tutors and argue that a pre-defined syllabus is beneficial for learning things that the student otherwise would not notice.

We can also differentiate CALL tools and CALL tutors using Vygotskian terminology introduced in subsection 2.1.1: while CALL tools only mediate learning, the aim of CALL tutors, even if not stated explicitly, is to actively guide students through their zone of proximal development. To do so, they should expose some degree of intelligent behaviour. Therefore, in this thesis we are concerned with the intersection of CALL and intelligent tutoring systems (ITS), which are described in the next section. It overlaps to a large degree with a sub-field of CALL known as ICALL (Intelligent CALL). We should note, however, that tutors may also give access to some CALL tools, such as a dictionary. Such integration enables the user to use the tool in an independent manner, while at the same time letting the tutoring system track how the user uses the tools in order to obtain information about the user's behaviour and knowledge. This information can be stored in the student model and subsequently used to adapt the tutor to the user's needs. A properly designed ICALL tutor can therefore combine many advantages of CALL tutors and CALL tools.

As noted in the introduction chapter, the field of CALL is more fragmented than the field of second language acquisition, with separate lines of research for teaching pronunciation, grammar, vocabulary and discourse. In subsection 2.1.5 we argued that vocabulary can be seen as a basis for many other skills. Therefore it will be the focus of the next subsection.

2.3.1 Vocabulary tutors

According to Ma & Kelly (2006, p. 18) vocabulary tutors can be divided into three broad categories: “multimedia packages with vocabulary learning activities”, “programs made up of written texts with electronic glosses” and “programs dedicated to vocabulary learning”. In this thesis, we will concern ourselves only with the last group. The integration of vocabulary tutoring with other activities, as in case of two other groups, may of course have advantages. However, as discussed in subsection 2.1.5, there are reasons to believe that learning vocabulary (as well as multi-word items) is an activity that can improve other language skills, and therefore vocabulary tutors are important as tools in their own right.

Ma & Kelly (2006, p. 16) describe two trends seen in modern vocabulary tutors: putting vocabulary learning into a wider context and giving “as much freedom as possible to choose what to learn and how to learn”. Wider context, be it situational or textual context, as long as it does not introduce confusion, offers clear advantages for vocabulary learning. The issue of giving as much choice as possible requires, however, more detailed discussion.

One problem is the possibility that giving too much choice may leave some learners confused about the most effective way of using the software.

Another issue with giving the user the choice of what to concentrate on, is the actual usefulness of vocabulary. The learner may know that some classes of words are not going to be useful for them at all (for example, it does not make sense to learn different names of fish if one cannot name them even in one's native language). The learner may, however, be easily confused about usefulness of a word due to different ways different languages split the word into concepts. For example, the Chinese–English CC-CEDICT dictionary defines the word 上火 *shànghuǒ* as “to have a yang 阳 imbalance in the body causing excessive internal heat” (“shànghuǒ, 上火” 2009). An intermediate learner might want to skip that item thinking it is a very obscure Chinese medicine-related technical term, while in fact, it is as everyday word as, for example, 感冒 *gǎnmào* “to have a cold”. We can conclude that leaving the learner too much choice about what needs to be learnt may easily lead to the ignoring of important concepts without direct equivalents, that is, ones that are harder to notice in the first place.

Of course this is not to say the user should be given no choice, but rather that the advantages and disadvantages of a particular solution should be considered, taking user's motivation and usefulness of choice into account. The tutor certainly should be able to suggest a reasonable teaching plan, if the user does not have any special preferences, but there should be an option to adapt the plan to the individual needs of the user. However, when it comes to repeating words that already have been learnt, asking the user what should be repeated serves little purpose. As long as the student model is fairly accurate, which we should expect from “intelligent” systems, we can find out what items will be most useful to repeat at a given time. In this case it is futile to let the user e.g. repeat items that do not need to be repeated, and it is better to make them focus on learning new words instead, if that is the most rational choice for a given situation.

After having discussed CALL tutors in general, and vocabulary tutors in particular, let us look at some principles for creating intelligent tutoring systems.

2.4 Intelligent Tutoring Systems

Intelligent tutoring systems (ITS) are computational agents whose purpose is to facilitate learning, usually without the help of a human teacher. Such systems may be designed for tutoring a particular area of knowledge, or be quite general. Examples of ITS include, among others, tutors of geography, mathematics, physics and computer programming. Substituting human teachers is of course not trivial, and therefore a research area concerned with building intelligent tutoring systems is sometimes called **artificial intelligence in education** (AIED). In this section we will provide general information about ITS and their structure.

We can distinguish three main models used in intelligent tutoring systems: the domain model, the student model and the tutor model

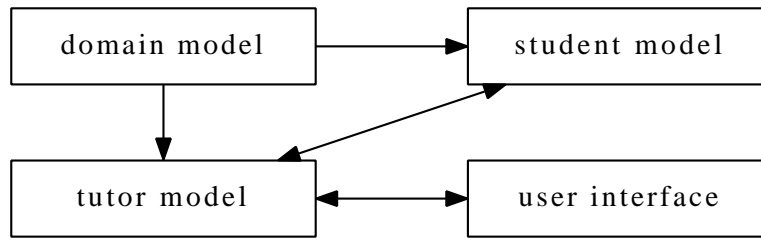


Figure 2.1: Typical relationship between models in intelligent tutoring systems

(Nkambou, Bourdeau & Mizoguchi 2010). Figure 2.1 presents a typical information flow between these models. The **domain model** represents knowledge that is to be taught by the system, and usually remains unchanged during the tutoring session. The **student model** contains information about the student. Most importantly, it tracks the student’s progress in gaining the domain knowledge, but it may also track other cognitive and affective states of the student that are relevant to learning. As opposed to the domain and student models, which predominantly store information, the role of the **tutor model** is to decide what should be done, taking into account the data about the domain, about the student and the tutoring module’s own internal state. The action is performed through the user interface, and the student’s response usually leads to updating the student model with new information, and selecting subsequent action by the tutor model.

The information flow presented in Figure 2.1 is a typical, but not the only possible option. For example, an ITS can, in principle, gather new information about the domain during the session, and update the domain model accordingly, or have asynchronous events in the user interface directly update the student model. Not only may the information flow vary. The three models are often interwoven to some degree in actual tutoring systems. Nevertheless, this component view is regarded as “classic” (Nkambou, Bourdeau & Mizoguchi 2010, p. 5) and remains important for structuring research in ITS. Therefore, in the following subsections the most important aspects of these models will be presented.

2.4.1 Domain model

The knowledge that the system aims to teach is represented in the domain model. Nkambou (2010, p. 17) makes a distinction between declarative (“knowing something”) and procedural knowledge (“knowing how to do something”) and notes that “most ITS have focused on procedural domains having limited scope”, and therefore there was not much need to pay attention to this distinction.

The noticing hypothesis discussed in section 2.1 may let us conclude that declarative knowledge generally precedes procedural knowledge in language learning: conscious attention allows us to discover the existence of new

language constructions (declarative knowledge), and only then may we start to internalise recognition and production of such constructions (procedural knowledge).

There are several general-purpose knowledge representation languages that can be used in intelligent tutoring systems to structure the domain model: production rules, semantic networks, conceptual graphs, frame-based systems, ontologies and description logics (Nkambou 2010).

A common feature of many representations is that they aim to express a useful subset of first-order logic. An advantage of a more restrictive language (semantic networks and description logics usually lack negation and disjunction) is making sure that inference is tractable, which is not the case in first-order logic in general (Russell & Norvig 2010).

Let us narrow our focus to **semantic networks**, which is a common term for various types of graph-based knowledge representation. Their common feature is that they contain **nodes**, usually graphically represented by ovals and **edges** (also called **links**), usually graphically represented by arrows. Nodes represent either objects or categories of objects, and edges represent relations between them. Edges are usually represented with arrows that indicate the direction of the relation.

Russell & Norvig (2010) point out that when linking categories with edges it is important to distinguish a relation between two categories from a relation between members of these categories. For example, the category *FemalePersons* may be linked to the category *Persons* with a relation *SubsetOf*, to indicate that the set of females is a subset of persons. On the other hand, when we make a link from *Persons* to *FemalePersons* labelled *HasMother*, we mean that every object that belongs to the category *Persons* is in relation *HasMother* with an object from category *FemalePersons*.

The domain model used in this thesis has a small number of categories. Therefore, we only need relations between objects, and even though they may belong to categories, such as constructions or exercises, there is no need to explicitly model relations between categories or organise them into a taxonomy. In subsection 3.3.1 in the next chapter we will present a such a domain model.

Modelling the domain with a semantic network gives us an advantage related to mapping it to a student model. We will see in subsection 2.4.2 that domain models are a sort of template for the structure of student models. We cannot observe directly what the user knows, so even if we represent our domain in terms of certain knowledge, it makes sense to have a probabilistic model of the user. The advantage of semantic network is therefore a straightforward mapping to a user model that represents uncertain knowledge, such as a Bayesian network, which will be used to represent the student model (cf. section 3.3.2 in the next chapter). As we have seen, semantic networks can help us to qualitatively present relations between objects. Bayesian networks will help us to quantify these relations. But before getting to that point we will make an overview of different possible student models and tutor models.

2.4.2 Student model

Finding better ways of modelling student knowledge in order to individualise system’s behaviour is seen as a big challenge for research in ITS in the near future (McCalla 2010). The growing importance of the Internet explains why this aspect of ITS is becoming increasingly more important. It is easy to use tutoring systems online from any place, which is probably one of the reasons why the issues of “life-long learning” are gaining importance. They involve modelling user’s knowledge over long periods of time, which requires more detailed student models than traditional single session-oriented ITS, and need to take into account issues such as forgetting.

As mentioned in the previous subsection, the student model often has a very similar structure to the domain model, since it is supposed to track how well the learner masters the domain knowledge. In the simplest case, it may be a checklist over elements of the domain model that marks which parts have been understood. More complicated models may track probabilities that something is known, and how they change over time. Since this may be seen as a transparent layer over the domain knowledge, such models are called **overlay student models** (Olney, Graesser & Person 2010).

More advanced student models extend beyond an overlay over the domain model. They may model student’s affective states, and, for example, find out when the student gets bored and let the tutor model find a strategy to increase their motivation.

Forgetting is another important aspect of learning that can be reflected in the student model. No knowledge can be retained forever without repetition, but teaching same items again during every session is definitely sub-optimal, therefore a model of forgetting is important for any tutoring system that is going to be used over long periods of time.

Woolf (2010) divides techniques for storing information in student models into two categories: methods originating in artificial intelligence (e.g. based on machine learning and plan recognition), and the ones coming from cognitive science (e.g. model-tracing and constraint-satisfaction).

Artificial intelligence–based methods

Machine learning techniques do not, in principle, make any assumption about the possibility of creating a psychologically plausible model of learning or mistakes, but rather try to generalise available student data. Such methods are not used in this thesis.

Probabilistic methods comprise another group of techniques that has its roots in AI. A model that is often used is a Bayesian network. In the case of tutoring systems, the core of such a network represents conditional probabilities among elements of students’ knowledge and between these elements and skills that the user might have demonstrated. Its versatility comes from the fact that it has an explicit way of expressing prior knowledge, with the structure of the network and prior probabilities, and a principled way of updating this knowledge, given new evidence. Woolf (2010) mentions two main ways of creating Bayesian networks: in **expert-centric** models

the conditional probabilities are based on human knowledge, while in **data-centric** networks they are learned from data.

Cognitive science-based methods

While cognitive science-based methods may take advantage of AI tools, which do not necessarily try to be cognitively plausible, their aim is to make the student model at some level functionally equivalent to a human learner and to base the structure of the model on theories of cognition (Woolf 2010). The levels of description may vary. We will now look at two important techniques that offer different approaches to modelling human learning: model tracing and constraint satisfaction.

Model tracing assumes that learning is essentially information processing. It is based on a theory of cognition called Adaptive Character of Thought – Rational (ACT-R). One of the most important assumptions of this theory is that human knowledge can be separated into procedural and declarative knowledge (Corbett & Anderson 1994). This distinction has been explained in subsection 2.4.1. The essence of this technique is that the system generates all possible actions that a system might perform in the current situation if it were the student, and tries to find the closest match to the action that the student actually performed. The aim of model tracing is to follow the user’s attempts at solving a problem and to understand what problem the user is facing at the given moment, in order to provide useful hints and feedback.

Another, related technique is **knowledge tracing** (Aleven 2010). Its goal is to model what the student knows. More specifically, it can be used to calculate the probability that the user has learnt a particular skill. Reye (1998) found that this algorithm can be implemented with a dynamic Bayesian network – a kind of Bayesian network that explicitly traces changes of probabilities over time (Russell & Norvig 2010, ch. 15). However, knowledge tracing in its original formulation has some important disadvantages. As Gong, Beck & Heffernan (2010) points out, it assumes that each exercise tests only one skill. Conati (2010) adds that knowledge tracing requires that each exercise have only one possible solution that the user must follow precisely step-by-step, without making any shortcuts. Moreover, there is no way of modelling dependencies between skills. Conati (2010) argues that more flexible, task-specific Bayesian network structures can be used to overcome these problems.

Other methods make weaker assumptions about the possibility of modelling human learning. Constraint-based modelling assumes that only errors can be effectively modelled, and concentrates only on the student’s mistakes, which are detected by checking that some constraints have been violated (Woolf 2010). The system does not model the knowledge of the student explicitly. There can be many solutions, but they are not listed explicitly. A set of constraints is given, and if they are met, we know that the user has given a correct answer. If they are not, the feedback is based on unmet constraints, without explicit modelling of what the user knows. This method is simpler than model tracing, but it still has a theoretical

motivation, as it is based on a theory of learning from performance errors (Ohlsson 1996).

Memory models

Memory models try to predict when something will be forgotten. An important point in quantitative research on human memory was a study performed by Ebbinghaus (1885/1913), that modelled the process of forgetting, based on the author's experiments on himself with remembering nonsense syllables. However, the idea of making an algorithm for effective learning consists of repeating facts in increasing intervals evolved quite independently of that. Mace (1968) suggested that effective learning should involve repetition with intervals increasing according to geometric series. Several modifications of such scheme have been applied in practice (e.g. by Sebastian Leitner and Paul Pimsleur). The first non-trivial spaced repetition algorithm and first computer implementation of such an algorithm is often attributed to Woźniak & Gorzelańczyk (1994). Woźniak devised the SuperMemo algorithm after a series of experiments similar to those by Ebbinghaus.

The memory models are related to cognitive science, as they aim at simulating an aspect of the human mind. In terms of methods, however, much of the above research has concentrated on generalising available data, rather than fitting a memory model into a theory of cognition. More recent research has, however, changed the focus. An example of an approach to memory modelling that aims at psychological plausibility is an extension to ACT-R developed by Pavlik (2006).

2.4.3 Tutor model

Bourdeau & Grandbastien (2010) contrast tutoring and teaching: teaching may be seen as a more general term, while tutoring is a specific form of teaching that involves high tutor/student ratio, high level of adaptation of tutor to the student, and high degree of interactivity. It has thus many advantages, the main disadvantage being the cost. ITS aim at providing these advantages at lower costs, tutoring is therefore often seen as a foundation for ITS.

Bourdeau & Grandbastien (2010) describe the growing understanding of the importance of interaction during tutoring within the ITS community. The **interaction hypothesis** has been formed, which states that effectiveness of tutoring increases with the degree of interactivity of the system.

One of the main directions of evolving ITS towards more intelligent behaviour is using **human language technology** (HLT) that allows more robust analysis of user input and more high-level rules in the program, and using HLT solutions, such as dialogue systems, which provide a high degree of interactivity, and allow the user to communicate in a natural language, which is the preferred way of communication for many users.

Bourdeau & Grandbastien (2010) point out that tutor models are in many cases tightly coupled to either domain or student models. The degree

to which they are intertwined may vary. Dubois et al. (2010) provide examples of different design of tutor models. They may for example use model tracing, which was mentioned in the previous subsection, and use *if-then* rules to decide on the next action, based on what the student did according to the model. The tutor model may also take form of a dialogue system. This approach will be further discussed in section 4.4. The system's knowledge about the tutoring situation can often be uncertain, and a principled way to deal with it, which has already been mentioned in the previous subsection, was to use Bayesian networks. Since tutoring is mainly concerned with making decisions, an extended version of Bayesian network, called decision network can be used to make rational decisions. These networks, known as probabilistic graphical models, will be discussed in section 3.2.

2.5 Summary

This chapter introduced important concepts related to learning. The **zone of proximal development** refers to the tasks that the learner cannot complete alone, but can complete with the help of a teacher or a more capable peer. The process of assisting the learner is called **scaffolding**. Research in scaffolding can give us an insight into the role of the teacher in facilitating learning.

There are many hypotheses about how people learn foreign languages. The **comprehensible input hypothesis** has important insight on the role of the input in language learning, and the **noticing hypothesis** explains why people sometimes seem not to learn despite constant exposure to comprehensible input. The **pushed output hypothesis** is related to the noticing hypothesis, and can offer some guidance about what learning strategies may be effective. The pushed output hypothesis is related to the issue of **corrective feedback**. The question of its effectiveness is controversial, but there are some forms of feedback that appear to be effective. Language teaching focused only on meaning cannot solve all learning problems and sometimes focus on form is needed. The **lexical approach** offers an interesting view on the nature of language that can be used to create principles of what should be taught in order to let the learners express themselves fluently and idiomatically.

There are many features of Chinese that are different from English. They include the use of characters, word segmentation ambiguity and use of classifiers. There is also no inflection in Chinese, and there are a very small number of grammatical markers. Chinese is an SVO language, but the basic word order is slightly different from English. In some cases, determined by use of specific words, the word order becomes more flexible. Aside from that, word order can be changed by topicalisation, which is more flexible than in English. Unlike English, Chinese often allows verb arguments to be omitted.

CALL applications include tutors and tools. A recent trend is concerned with building tutors that expose some degree of intelligent behaviour. There

are reasons to believe that tutors specifically designed to teach vocabulary, that intelligently adapt curriculum to every individual, can be beneficial not only to extending vocabulary, but also to other language skills.

Intelligent tutoring systems typically consist of a domain model, student model and tutor model. The **domain model** represents the knowledge that the system aims to teach. It can be represented as, among others, a semantic network. The **student model** collects various information about the user. The most important thing is to model what the student knows. Also memory is an important factor: the student model should know when some knowledge is going to be forgotten and should be repeated. Student models can be based on AI or cognitive science models. Bayesian networks are important tools used in student models. The **tutor models** is concerned with making decisions on what the tutoring system should do. Recent research stresses the importance of interactivity during tutoring. The student model is the most important source of information for the tutor model. Decision networks are particularly suited to make tutoring decision for systems with Bayesian student models.

Chapter 3

Approach

This chapter presents the approach to designing an intelligent tutoring system for learning Chinese. It is based on the theories presented in the previous chapter. The core of the approach is in the cognitive modelling of the learner's vocabulary knowledge and in selecting the next exercise given available information. This is the main contribution of this thesis. The implementation of the approach will be presented in the next chapter, and its empirical evaluation – in chapter 5.

Section 3.1 discusses rationale and the general design of the system. Section 3.2 introduces probabilistic graphical models, the tool used at the core of the approach. Section 3.3 describes how these models are used to represent knowledge in the system, in the form of a Bayesian network, and to select the next exercise, using a decision network.

3.1 Rationale and General Design

Consider the task of creating a tutoring system for teaching vocabulary. As discussed in section 2.1, pushing learners to produce sentences is a good way to make them notice new constructions, that is, to get some declarative knowledge about their existence (cf. subsection 2.4.1). Making it into procedural knowledge requires repetition. The optimal level of such sentences is determined by student's zone of proximal development (cf. subsection 2.1.1).

3.1.1 Exercise types

The problem with most vocabulary tutors is that they usually provide the kind of language exercises that are found in language textbooks, such as finding synonyms, grouping the words based on meaning or filling in the gaps. They are not clearly related to the language activities of a user in natural situations. The exercises in which a user would construct a sentence with a particular meaning are not suited to traditional textbooks – there may be a large number of possible answers and it is not feasible to list them all.

There are tutoring systems that use natural language processing

techniques to analyse sentences provided by the user and give feedback (see e.g. Heift & Nicholson 2001). However, the focus of these systems is usually on grammar. They may provide a range of vocabulary-related feedback, e.g. about choice of the words in collocations, but teaching vocabulary items is not their main goal.

We argue that making the user construct whole sentences and providing feedback is a method vocabulary tutors can employ, too. An obvious advantage is that creating sentences is what the learner ultimately does in natural settings. In real life situations people want to express a particular meaning, relevant to the actual context, and they use the vocabulary and grammar they know in order to express that meaning. There are no sentences with gaps to fill, nor words to group.

Another advantage of letting the user write whole sentences is that it gives more information to the tutor. The user's vocabulary knowledge is not directly observable; the structure of the exercises should therefore allow the collecting of as much evidence as possible about what the user knows. Fill-in-the gap exercises test the knowledge of one particular word or word group, and assume that all the other words in the sentence are known. The system gets evidence about whether the word that fits into the gap is known, but there is no direct evidence about knowledge of all other words in the sentence. Successful completion of the exercise makes knowledge of all the words in the sentence more likely, but it cannot be regarded as strong evidence. If the learner writes whole sentences, the system gets much more information – it can collect evidence about knowledge of individual words, by checking whether they appear in the sentences written by the learner or not. With more information, the system can more easily determine the user's vocabulary knowledge and provide more useful exercises.

There are many ways of making the learner produce sentences in the target language in the context of an ITS. A quite natural approach would be to make exercises that introduce some realistic situations (e.g. shopping, looking for a job, discussion about some topic), and make the user interact with interlocutors played by the system. Such approach has, however, some disadvantages. An important one is the relative difficulty of constructing exercises, as they have to account for quite many possibilities in the interaction.

Another possible approach, taken by de Vries et al. (2014), is easier to model than a full dialogue. They make the system introduce some context (e.g. a story) and ask questions about it. However, this approach may not work well for a vocabulary tutor. Introducing a story takes some time, so each exercise takes a relatively long time to complete (when the context is introduced before each question) or the exercises are less fine-grained (when several pre-defined questions are always associated with one exercise). In the previous chapter we argued that an intelligent tutoring system should be able to choose exercises, based on what it knows about the user. Smaller and more constrained exercises give more opportunities for the system to schedule the exercises, and, if the scheduling algorithm works well, lead to more effective learning.

3.1.2 Translation exercises

A simple type of exercise that does not have the above-mentioned problems, is the translation exercise: we can provide a sentence in a language the user already knows, and ask for a translation. The number of correct translations may be large, but as we will see in the next chapter, it can usually be encoded in a compact manner. It is definitely easier than encoding whole context of a dialogue, which is a non-trivial and time-consuming task that would be required if we wanted the exercises to provide full-fledged, natural interactions. Moreover, if the sentences to translate are properly chosen, they give enough information about their context to allow an unambiguous translation. Therefore, the translation exercises were chosen as the building blocks of the presented tutoring system.

There may be various degrees of acceptability of a translation. It is relatively easy to reject a translation when it violates explicit rules of grammar. The lexical approach, briefly introduced in subsection 2.1.5, emphasises the fact that a grammatical sentence may be unidiomatic. Native speakers of a language can usually agree to a certain extent about whether certain phrasing is idiomatic. However, when assessing a translation of a particular sentence on the level of semantics and pragmatics, the acceptability judgements of different native speakers may vary significantly: a slight difference in meaning may be acceptable to some and not to others, some people may be less used to a formal register, etc.

In our approach, we use the acceptability judgements of a small group of native Chinese speakers to formulate the answer sets for the exercises. They may not be representative for all, or even the majority of Chinese speakers. This is, however, not a problem specific to computer-based tutoring. In traditional scenarios that involve a human tutor, the tutor also uses their personal judgement to decide whether to accept the student's utterance or give negative feedback.

A sentence in one language can have many translations into another language. There can be two different reasons for that: either the sentence may be ambiguous, and have several different interpretations that lead to different meanings; or there can be different ways of formulate the same meaning in the target language. We want to avoid the first case: the aim is to elicit a target language sentence with a particular meaning, therefore we want to make sure that the source sentence is not ambiguous. The lack of ambiguity does not have to be absolute; the words that are polysemous in the same way in both languages do not pose a problem. The second reason for multiple translations is unavoidable: given an unambiguous sentence in one language, one can usually produce several equivalent translations in another language. Those several equivalent translations may be completely different from each other; however, there is usually some similarity between them. In the case of Chinese, we can identify two main ways in which translations of the same sentence can differ: they may use different synonyms or the word order may be different.

Some Chinese synonyms can be used interchangeably to a large degree, others may have similar meaning, but differ in the actual use. Depending

on the target group of the exercises, we may make additional simplifying assumptions. For example, exercises at the beginner and lower-intermediate level need to take frequently occurring synonyms into account, but do not necessary need to accept the whole range of rare synonyms that only occur in literary works.

As for the word order, it is not necessary to take every possibility into account. In subsection 2.2.4 we have seen various reasons for variable word order in Chinese. Topicalisation makes the sentence marked. One can argue that in most situations, a sentence in the neutral word order in one language should be translated into a sentence in the neutral word order in another language. When the users translate it into a sentence with a marked word order, it is likely that they are unaware of the markedness. Therefore, from a pedagogical perspective, not including the topicalised sentences among the correct answers should be a good thing. On the other hand, changing the order of some co-verbs, such as 给 *gěi* (discussed in subsection 2.2.4), does not make the sentence more marked. Therefore, such instances of flexible word order need to be taken into account.

3.1.3 Feedback

The notion of feedback, which is an important part of the pushed output hypothesis, has been shortly discussed in section 2.1. In this thesis we are not interested in forms of feedback that appear in actual conversations, but rather in the information it conveys. Language tutors usually employ explicit feedback. As mentioned in subsection 2.1.4, it is focused on form, that is, it is issued not only when the utterance is not understandable, but also when it is understandable, but ungrammatical (or, perhaps, unidiomatic). It happens rarely in natural situations, but is often used in language teaching contexts. It helps the learner to construct a correct sentence, and notice right and wrong ways of using grammatical and lexical constructions. The positive feedback does not need to do anything more than confirm that the sentence is correct, and strictly speaking, it is not necessary, but it may lead to more noticing, and reinforcing correct ways of saying things.

We also want to fulfil the principles of the dynamic corrective feedback, which is likely to be effective (cf. subsection 2.1.4): the system should provide it always directly after the user inputs a sentence (make it timely and constant), give information about one error at a time (so as not to overwhelm the user and make the feedback manageable), and provide enough information to correct the error (make it meaningful). When possible, we also want to use the prompt feedback, which is likely to be effective in interactive contexts: for example, when a word is missing, the system should not provide the Chinese word that can be directly put in the sentence, but rather provide an English word, and let the user find the Chinese word. The user may know the Chinese word, or may need to look it up in a dictionary. In the next subsection we will look at how a dictionary may be integrated with the system.

3.1.4 Evidence

We want the system to collect evidence about a user's knowledge. The most obvious way of getting evidence is the results of exercises. A successfully completed exercise (that is, a sentence correctly translated into Chinese) indicates that the user knows all the words and constructions that were used to construct the sentence. Of course, such evidence is not completely sure. The user might not have known some of the words, and might have looked them up in a dictionary. If the system has no information about that, the fact that the user completed the exercise carries less information.

It is therefore useful to integrate a dictionary into the system. As discussed in section 2.3, integrating CALL tools into tutors may have benefits. A dictionary mediates learning and extends the students' zone of proximal development: they can finish exercises that would be too hard without access to a dictionary. Moreover, offering the possibility to easily look up words during the exercise gives the user more choice. In order to make the situation realistic, we do not want to simply provide translations of words that are used in the exercises, but rather give access to a full dictionary, with many different translations, which requires the user to choose the one that is correct in a given context. This promotes users' independence and prevents them from using a dictionary in a mechanical way. Still, it may lead to over-reliance on a dictionary in a longer perspective. It is therefore important for the system to consider dictionary look-ups as further evidence of user's knowledge. If an exercise has been finished successfully, but a word has been looked up, it means that the word has not been learnt yet, and requires more repetition. Only after an exercise is finished without looking up that word, does the system get evidence that the word has been learnt.

We cannot observe users' knowledge directly. We want the system to get information about it from two main sources: exercise results and dictionary look-ups. Neither of them can give us certain information. Users may, for example, look up words that they know perfectly, or may skip some exercises, even though they are able to answer them. There are many unpredictable reasons for such behaviour. Moreover, we not only want to reason about user's knowledge, we also want to use this information to select the most beneficial exercises. It is difficult to manually account for all possible evidence, and create reliable and consistent rules for selecting exercises. We need a principled approach for reasoning and making decisions under uncertainty. This can be done with probabilistic graphical models, which are described in the next section.

3.2 Directed Probabilistic Graphical Models

In this section we present models that can represent and quantify uncertain knowledge, and use it to make decisions. They are represented with graphs, and are called **probabilistic graphical models** (Koller & Friedman 2009). Below we will present two PGMs that are represented with directed graphs: Bayesian networks and decision networks.

3.2.1 Representation of Bayesian networks

A **Bayesian network** (Pearl 1988) is a tool for representing a joint distribution of random variables. Each variable in consideration is represented as a single node in a graph, and edges represent their conditional dependence. For each variable X , all the variables Y_i connected with X by a directed edge that goes from Y_i to X are called **parent variables**. For each node of a Bayesian network, there is a conditional distribution of the associated variable, given its parents: $P(X|Y_1, \dots, Y_n)$.

If we consider discrete variables, Bayesian networks contain the same information as a table that assigns some probability to every possible combination of variables. They have several advantages over such direct representations. First of all, they allow for efficient inference algorithms – when we directly observe some of the variables in the network, we can compute the probability of some other variables in the network, given that knowledge. Moreover, when n events are not independent, there are 2^n possibilities to consider – each of them could have occurred or not, and we need to represent a probability of each such possibility. Bayesian networks utilise the fact that in most cases we can find groups of variables that are independent of each other, as long as we know something about another variable (conditional independence).

There are many ways of representing the same joint distribution by using different node ordering, which determines the edges between nodes. It is generally advantageous to choose a representation that makes best use of conditional independence (and thus has as little edges as possible) and facilitates knowledge engineering. Knowledge engineering is easiest when edges represent the sort of relations that lead to natural probability judgements. The same joint distribution can be represented in a model with arrows in any direction. Russell & Norvig (2010) argue that people usually have fewer problems with judging probability when the edges go from causes to results. Moreover, in a fine-grained model, events may have very many direct consequences, but they usually do not have many direct causes. With the opposite directions of the edges, each cause depends on its consequences, and we need to consider likelihood of the cause given each combination of the consequences.

There are two kinds of random variables that can be represented in Bayesian networks: discrete and continuous variables. We need a way to find out the distribution of value of the variable, given the values of nodes connected by incoming edges (**parent nodes**). In this thesis we will mostly use discrete variables. The conditional distribution of a discrete variable that depends on other discrete variables can be described in the form of a **conditional probability table** (CPT). For each combination of values of parent nodes (a **conditioning case**), the table specifies a probability distribution of a variable. CPT cannot be used directly in the case of continuous variables.

Consider a node without parents. If it is a discrete variable, listing all possible values of the node and their probabilities (that need to add up to 1) is a sufficient description of the probability distribution. If it is a continuous

variable, we cannot enumerate all its values. The most common solution to this problem is assuming that the variable is distributed according to some known probability distribution (e.g. a normal distribution), which can be described using a finite number of parameters. There are also other ways to solve this problem, without using parameters (non-parametric representations), but they will not be used here.

Exact inference that involves both discrete and continuous variables might not give a closed-form solution. A simple way to overcome this problem in Bayesian networks is to use **discretisation**. In a discretised representation of a continuous variable, the range of its possible values is divided into intervals. Each interval is associated with the probability that the variable has value in this interval. An advantage of this approach is simplicity – inference and learning procedures used for discrete nodes can be used in nodes that originally were continuous. However, this method also has disadvantages: the conditional probability table can be large (if we use a large number of intervals) or the inference can be considerably less accurate (with a small number of intervals).

Let us now look at a simple example of a Bayesian network, presented in Figure 3.1, that can model the probability of passing a particular language exam. We have three random variables: *Pass the exam* (PtE), indicating whether a person passed the exam, *Learning time* (LT), indicating how many hours a person has been learning that language, and *Majority language* (ML), indicating whether a person lives in a country where that language is spoken by majority of inhabitants. *ML* and *PtE* are clearly discrete variables, which can take two values: *true* or *false*. The most natural way to represent *LT* would be a continuous variable. In this case, however, we used discretisation, and divided the continuous range of values into a discrete set of sub-ranges. This means ignoring any differences in values the variable can take within each range.

3.2.2 Inference in Bayesian networks

The inference in Bayesian networks consists of a sequence of two kinds of operations: marginalisation and applying Bayes' rule. Consider a task of computing probability of passing the exam, given that a person spent between 50 and 100 hours learning the language. Under such assumptions, we can ignore all the rows in the CPT for *PtE*, apart from the third and fourth. We do not know what the value of *ML* is, we know, however, that it is *true* with probability 0.2, and *false* with probability 0.8. Therefore we need to consider both options, and sum them out, weighted by their probability. This operation is called **marginalisation**:

$$\begin{aligned} P(PtE|LT \in [50,100]) &= P(PtE|LT \in [50,100], ML)P(ML) \\ &\quad + P(PtE|LT \in [50,100], \neg ML)P(\neg ML) \\ &= 0.5 \times 0.2 + 0.3 \times 0.8 = 0.34 \end{aligned}$$

Now if we want to ask what the probability that a person spent between 50 and 100 hours learning the language is, given that they passed the exam,

Learning time			
0-50	50-100	100-500	500+
0.3	0.4	0.2	0.1

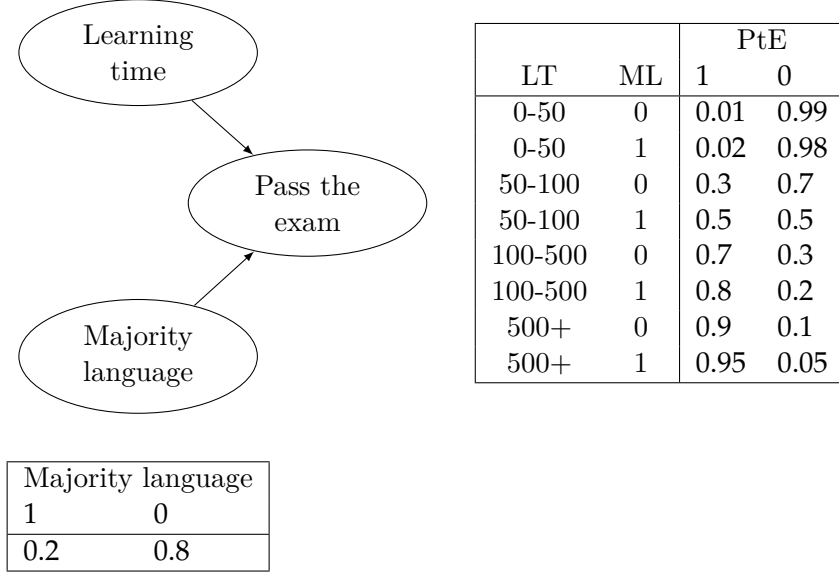


Figure 3.1: Example Bayesian network

we need to use the **Bayes' rule**.

$$\begin{aligned}
 P(LT \in [50, 100] | PtE) &= \frac{P(PtE | LT \in [50, 100])P(LT \in [50, 100])}{P(PtE)} \\
 &= \alpha P(PtE | LT \in [50, 100])P(LT \in [50, 100])
 \end{aligned}$$

We can read the values directly from the CPT, apart from the value of α , which is a normalisation factor that replaced $\frac{1}{P(PtE)}$. Since the values of LT exhaust all the possible cases and exactly one of them is true in every possible world, they must sum up to 1:

$$\begin{aligned}
 &P(LT \in [0, 50] | PtE) + P(LT \in [50, 100] | PtE) + \\
 &P(LT \in [100, 500] | PtE) + P(LT \in [500+] | PtE) = 1
 \end{aligned}$$

We can compute each addend in the way described above, and use the equation to find the value of α .

3.2.3 Noisy functional dependence

The example shows that even very simple Bayesian networks may have relatively big CPTs. The bottom table in Figure 3.1 already has 8 rows, every new parent of PtE would double the number of rows. In a larger network the number of parameters may quickly become unmanageable. One possible solution is to reverse the relationships. If we do that, PtE will have

no parents, and it itself will be the only parent of other nodes. Any node that would be a parent of *PtE* in the original structure, would become a child of *PtE* here, and would not cause the exponential growth of CPT. Any distribution that can be represented in the original structure, can also be represented in the new one. However, as we have already mentioned, such a structure needs to make unnatural probability estimates. Causality goes from *LT* to *PtE*, and not the other way round. We will therefore look at another possibility of reducing complexity of CPT: noisy functional dependence.

Some causal relationships between parent nodes and a child node resemble the logical OR relation. For example, if two parent nodes are *Sprinkler turned on* and *Rain* and the child node is *Grass is wet*, a natural CPT for the child node will give it high probability when any of the parents are true – the grass may be wet because of the sprinkler, the rain or both. On the other hand, if the parent nodes are: *Intercourse* and *Contraceptives*, and the child node is *Pregnancy* (a simplified example taken from Díez & Druzdzel 2006), a natural CPT for the child node will give it high probability only in one case: when *Intercourse* is true and *Contraceptives* is false – in other cases the pregnancy is unlikely. It is easy to see why a function that can provide CPT values in the first case is called noisy-OR, and the second one is called noisy-AND. Only the latter will be used in this thesis.

The noisy-AND model assumes that each of the conditions (represented by parent nodes) may be inhibited or substituted (Díez & Druzdzel 2006). If a node that is true is inhibited, the result is false, even if all other conditions are true. If a node that is false is substituted, it does not cause the result to be false, as it normally would. The model requires two parameters for each parent of the node in question: c_i , which represents the probability that the node i is not inhibited, and s_i , which represents the probability that the node i is substituted.

Díez & Druzdzel (2006) provide the following formula for calculating the noisy-AND CPT for a child node y and parent nodes x_1, \dots, x_n :

$$P(y|x_1, \dots, x_n) = \prod_{i \in I_+} c_i \prod_{j \in I_-} s_j$$

where I_+ is the set of parents that are true, and I_- is the set of parents that are false.

3.2.4 Utility theory and decision networks

Utility theory lets us model rational decisions. Its main principle says that rational agents should maximise their expected utility. There is a set of actions that an agent may perform at a given time, which may lead to some outcomes. We assign a numerical value (**utility**) to each of these outcomes. The expected utility of an action a is the sum of the utilities of the possible outcomes of that action, weighted by the probabilities of each outcome. According to the utility theory, as long as we have some consistent preferences among the outcomes, we can always find an assignment of

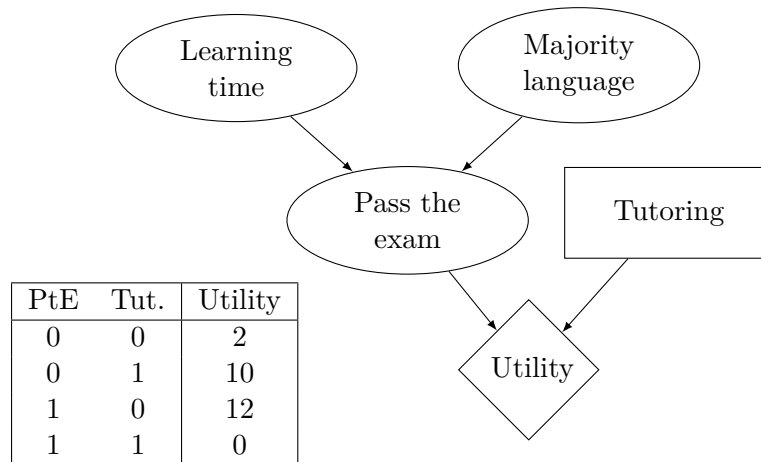


Figure 3.2: Example decision network

numerical utilities to outcomes so that maximising such expected utility will always be a rational decision (Russell & Norvig 2010).

Decision networks, also called **influence diagrams** (Pearl 1988) combine the knowledge stored in a Bayesian network with an utility function and calculate which decision leads to maximising the expected utility. They contain **chance nodes**, which are the same as nodes in a Bayesian network. Additionally they contain **decision nodes** and **utility nodes**. Decision nodes take part in inference in the same way as chance nodes, but they do not have any incoming edges, because they are not random variables, but potential decisions that are considered. Utility nodes, on the other hand, have only incoming edges from nodes that are relevant for calculating utility.

In the decision process, for each potential decision that can be made at a given time, we calculate the value of random variables, given that decision, which are in turn used to calculate expected utility of the decision. Then, a decision with highest expected utility can be made.

In the original formulation, the utility function depends on a utility and probability of possible future states caused by an action. In some cases, however, it is easier to directly estimate expected utility of an action, given some states, in form of an **action-utility function**. In other words, instead of asking “what is the average utility of the outcome of action a ?”, we can ask “given that we are in the state s , what is the utility of action a ?”.

Figure 3.2 presents an example decision network that can be used for deciding whether one should pay for additional expensive tutoring to pass the exam. The chance nodes are represented as ovals, decision nodes as rectangles and the utility nodes are diamond-shaped.

Tutoring should obviously positively influence the chances of passing the exam. However, instead of estimating the probability of passing the exam given tutoring, we simply took this relationship into account when creating the action-utility function presented in the table. The fact that tutoring is expensive is not directly represented on the diagram either. It is reflected in

the utility values: not taking tutoring has on average higher utility values than taking tutoring.

The table was constructed under the assumption that tutoring has a strong positive effect on the possibility of passing the exam. If we knew for sure that given the student's current knowledge, the exam cannot be passed, the tutoring would be certainly useful. Therefore, in the first two rows of the table, the utility of tutoring is higher than the utility of no tutoring. Conversely, if we knew for sure that given the student's current knowledge, the exam will certainly be passed, there would be no point in the tutoring. Therefore, in the last two rows of the table, the utility of no tutoring is higher than the utility of tutoring.

If we know that given all the available evidence, the probability of passing the exam is 0.34, we can compute the expected utility (EU) of tutoring (T):

$$EU(T) = P(PtE)U(PtE, T) + P(\neg PtE)U(\neg PtE, T) = 6.6$$

Analogously, we compute the expected utility of no tutoring:

$$EU(\neg T) = P(PtE)U(PtE, \neg T) + P(\neg PtE)U(\neg PtE, \neg T) = 5.4$$

The utility of taking additional tutoring is higher in this case; therefore, if the utilities accurately represent our preferences, this is the rational decision to make.

3.3 Formalisation

This section presents a formalisation of the approach, using the tools presented earlier. We will now define the three models found in intelligent tutoring systems. First we present the domain model, which defines what we want to teach. Then we discuss the structure of the student model, which models the knowledge of the learner. Finally, the tutor model makes decisions: based on the information about the learner's knowledge stored in the student model, it selects an exercise that is most likely to benefit the user.

3.3.1 Domain model: Constructions in sentences

Constructions

Let us assume that we have a list of translation exercises, and that each exercise contains an English sentence and a list of equivalent Chinese sentences. The Chinese sentences may be represented as patterns; each pattern may contain actual Chinese words (which we will call **terminals**), but may also contain special symbols (**non-terminals**) that indicate a group of synonyms that can be used in a given context. The actual exercises that were created according to these principles are presented in the appendices.

From such exercises, we can extract Chinese-English equivalences: a group of Chinese words that are synonymous in the context of a given exercise can be linked to an English phrase that describes its meaning. The

words themselves do not need to have an English translation. For example, a list of two words (计算机 *jìsuànjī*, 电脑 *diànnǎo*) can be linked to an English translation “computer”, while a list of words (栋 *dòng*, 座 *zuò*) can be linked to an English description “classifier for buildings”, which describes the function of these words. We will call such pairings **constructions**. We can note that the English translation/description can be regarded as the meaning of the construction, and the list of Chinese words is a list of forms that the construction may take. We use **construction** as a term for mapping between form and function – it is used in this way in the Construction Grammar (Goldberg 2006).

We can assign unambiguous names to constructions. For instance, the construction that is described above as “classifier for buildings” can be called [m.buildings]. Now this construction can be used in any exercise where either 栋 *dòng* or 座 *zuò* can be used as a classifier for buildings in the correct answer. However, these two words are not interchangeable in all contexts. If there is an exercise where, for example, only 栋 *dòng* can be used, we cannot use the construction [m.buildings], but we need to create another one that contains 栋 *dòng* and possibly some other words and has another name.

Multi-word constructions

So far we have discussed constructions that refer to lists of individual lexical items, but constructions can be larger. Consider for example, the conditional construction “if..., then...”, which can be expressed in Chinese in different ways, for example (words in brackets are optional):

[如果]	...	[的话]	,	[所以]	...	就	...
<i>rúguǒ</i>		<i>dehuà</i>	,	<i>suǒyǐ</i>		<i>jiù</i>	
if		if		then		then	

We can treat this whole construction as one item. It may contain as many as four words with several placeholders to fill (or may just consist of one word 就 *jiù*), but it expresses one clear meaning – conditionality. It is therefore useful to track the learning of this construction as a whole. It is a mapping between some forms and a meaning, and can therefore be regarded as a construction. Such uniform treatment of individual words and larger constructions is therefore similar to the approach taken by the Construction Grammar.

This is also consistent with the lexical approach to language teaching that was mentioned in subsection 2.1.5. It maintains that the major part of learning a language consists of learning lexical items, understood not just as words and collocations, but basically any non-random chunks that appear in actual language use, possibly with some placeholders where different words can be placed.

Learning targets

The teaching domain consists of constructions that appear in the exercises. In this subsection we will define what we want to become the **learning**

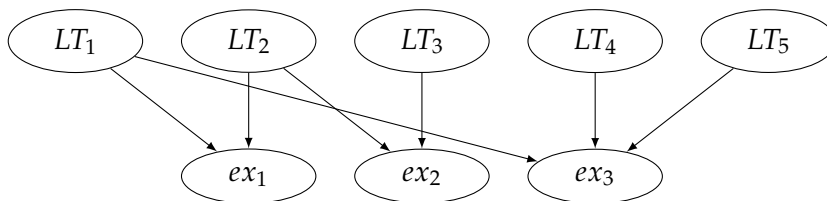


Figure 3.3: Example domain model represented by a semantic network

targets, that is, the items that the system will actively try to teach.

For a given set of translation exercises, it is not practical to use all the constructions that appear in it as learning targets. We are not concerned with teaching the most frequent vocabulary items, such as e.g. personal pronouns. Translation exercises are probably not a good way of starting to learn if somebody has absolutely zero knowledge of the language. We assume therefore some minimal command of language: that the user knows at least a few tens of the most basic words and can use them to construct simple sentences.

Constructions that appear only once in the whole exercise set are not good candidates for learning targets either. Using a word in only one context is usually not enough to internalise it; we want therefore words to be repeated across many exercises.

Representation

We have described the principles for choosing constructions from the exercises as the learning targets. The domain model needs to represent the link between the learning targets and the exercises. Semantic networks, which were described in subsection 2.4.1, can provide a natural representation for such relationship. Moreover, they can be extended to describe relationships between the learning targets. Semantic networks are well-suited to such applications: they are used to build lexical databases that are structured on psycholinguistic principles, such as WordNet (Beckwith et al. 1991).

Figure 3.3 presents a simple example domain model. There are five learning targets and three exercises. The edges indicate which learning targets appear in which exercises. They may also be used to indicate relations between the learning targets that are relevant to learning. For example, we might include edges that indicate that a learning target consisting of an individual word appears in another learning target that consists of a multi-word construction, or edges that indicate relations such as synonymy or antonymy. In this example, however, we assume that the learning targets are independent of each other.

3.3.2 Student model: Probabilistic information about learner’s knowledge

The system’s student model is the core of this thesis. Its aim is to represent the student’s knowledge in order to determine the student’s zone of proximal development (cf. subsection 2.1.1) and find out what exercise will be most beneficial to the student’s learning at any given moment.

In subsection 2.4.2 we presented an overview of student models, and noted that they often form an overlay over a domain model. We will therefore take the domain model presented in the previous subsection as the starting point, make an overlay that represents student’s knowledge, and then extend it further.

As already noted in this chapter, the interaction with the program cannot give certain evidence about what the user knows, and therefore we need a probabilistic model, such as a Bayesian network. We make a simplifying assumption that words and constructions selected as learning targets are the only parameters that determine the difficulty of a sentence. We also assume that the learning targets can be in one of only two states in the learner’s mind: *known* and *unknown*, and the probability is used to express incomplete knowledge about these states. We can now structure the core of the Bayesian network after the domain model. The edges can indicate the direction of influence: knowing all the learning targets that appear in an exercise has a positive influence on the probability of finishing the exercise. The information about the student’s vocabulary knowledge can be represented in the conditional probability tables (cf. subsection 3.2.1) of the nodes representing learning targets.

The network has a structure that may cause problems: there can be many learning targets in an exercise, and therefore an exercise node may have a large number of parents. We discussed this issue in subsection 3.2.3, and noted that a common solution to such a problem is to use noisy functional dependence. An exercise is likely to be successfully completed only when the user knows all the learning targets that appear in it. It is therefore a natural setting for using the noisy-AND function to create the CPT. One might try to differentiate the inhibition and substitution probabilities for different types of learning targets, depending for example on their word class or frequency. We will, however, make the assumption that all learning targets in all exercises share the same inhibition and substitution probabilities.

The learning target nodes are **hidden** – we cannot directly observe their values. We can, however, observe the user doing the exercises. Most importantly, we know the outcome of an exercise, whether it was successfully completed or not. We can therefore use Bayesian inference to modify CPT of the hidden nodes as the system gets information about exercise outcomes. The learning outcome can provide quite strong evidence about what the user knows: a successfully completed exercise implies that the user created a sentence with the learning targets related to the exercise, and they are more likely to be known. The lack of success, however, does not indicate which of the learning targets was problematic.

We can, however, find more evidence about the user’s vocabulary

knowledge. As mentioned in subsection 3.1.4, the system can be integrated with a dictionary. A word look-up is an indication that the user does not know a particular word. Since we are concerned with English-to-Chinese translation exercises, the user will most likely look up English words in order to find their Chinese translations. The learning targets need therefore to be associated with lists of English words whose look-up indicates that the learning target is unknown. In order to integrate the evidence obtained from look-ups, we need to create additional nodes in the Bayesian network. We assume that several look-ups of the same word during one exercise count as one, but looking up the same word in different exercises are separate pieces of evidence. Therefore, for each learning target node, we will create look-up nodes, one for each exercise that contains this learning target.

We need to consider the question of prior probability, that is, the assumptions about what the user knows before the system gets any evidence. We could make all the learning targets have 50% probability of being known, but this would not be efficient. For most users, the system would have to separately collect evidence about each individual word. Learners do not learn words randomly, but they do it in a more or less predictable order. Therefore, we can divide learners into levels. While the system can never be sure whether a user at a particular level knows a particular word, some words are definitely more likely to be known by users at some levels, and unknown by users at other, lower levels.

Such defined levels are not directly observable, and we need an indirect way to assess them. We will use two methods: users' self-assessments about their level of written and spoken Chinese, and a character recognition test. For languages such as English, the ability to tell real written words from made-up ones can be used to estimate the learner's vocabulary, and indirectly indicate their general written proficiency. In the case of Chinese, there is no unambiguous definition of word, and therefore we will use characters instead, even though they usually represent morphemes rather than words. We can assume that self-assessments and the results of the test will be influenced by the user's actual level. This, again, lets us use Bayesian inference to use observable data to assess the user's level.

Figure 3.4 presents the structure of a student model with three exercises and five learning targets. White ovals represent hidden nodes, and grey ovals represent observable nodes. Note that *number of chars* is a continuous variable, but in order to simplify the inference we may discretise it. The model can be used as follows. The system asks the user to make a self-assessment of written and spoken proficiency in Chinese. Then it asks the user to do a character recognition test, which provides an estimate of the number of character they know. Based on these three values the system performs the inference and calculates the probability distribution of the user's level. Based on this distribution, for each learning target, it calculates the probability that the user knows that construction. During the session with the program, the evidence about look-ups and exercise outcomes becomes available, and is used to infer new probabilities that the user knows the learning targets.

In order to make the inference, the network needs to have CPTs. In

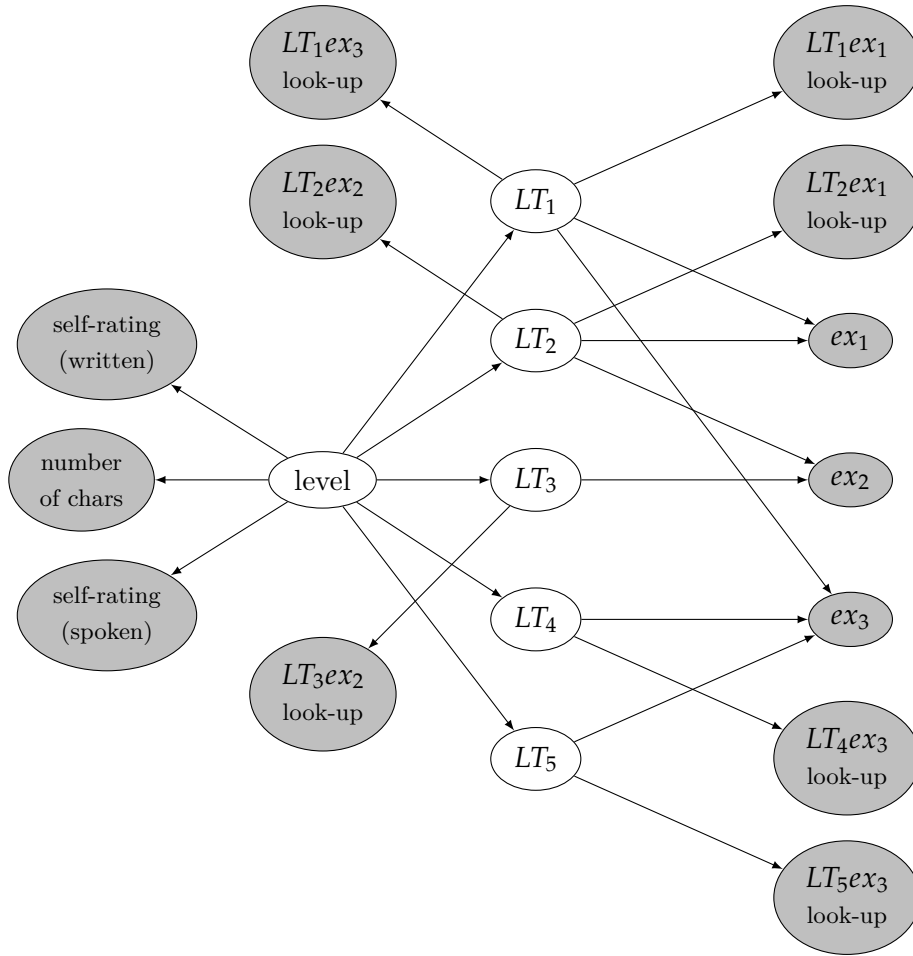


Figure 3.4: Example student model represented by a Bayesian network

chapter 5 we will present how they have been estimated. The aim of the student model is to provide data that can be used to select the next exercise. Next subsection provides more information about how this is done.

3.3.3 Tutor model: Selecting the next exercise

The main advantage of using a cognitive model to select exercises is that such a model should be able to find exercises at the right level. The information about what the user knows is used to select an exercise that is most likely to maximise the learning outcome at the next time step. It can be done by selecting sentences that are most likely to lie within the zone of proximal development. The ZPD consists here of sentences that the user would not be able to translate without any help, but could translate given system hints and access to the dictionary. The selected sentence cannot be too easy (if everything is known, the user will not learn anything), nor too difficult (if almost every word is unknown, the user will probably be too confused to finish the sentence successfully, and even if they do, it is unlikely they will remember the words).

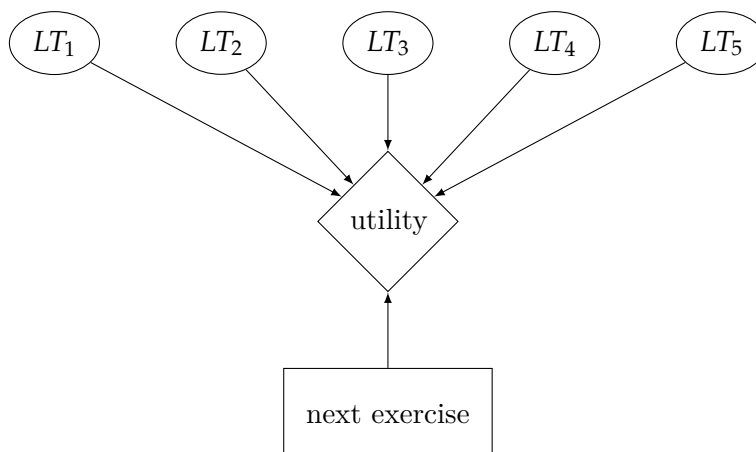


Figure 3.5: Example tutor model represented by a decision network

The next sentence is chosen by a decision network. This allows us to express what sentences we prefer, given the user’s knowledge. The structure of the network is presented in Figure 3.5. The utility function depends on the known learning targets and the choice of an exercise. The function needs to assign the lowest utility to sentences with all known learning targets, and the highest utility to sentences in which some, but not too many, learning targets are unknown. The problem of choosing the utility function will be discussed in more detail in chapter 5.

3.4 Summary

This chapter presented the approach to designing an intelligent tutoring system for learning Chinese, giving much focus to how a cognitive student model can be created. The basic building block of the system is an exercise. Translation exercises are useful for the learners and relatively easy to create. The system should push the user to write sentences slightly above their current level, and provide dynamic corrective feedback.

The problem of choosing the optimal level of the exercises is closely related to the student modelling. Our approach is based on probabilistic graphical models: the Bayesian network, which can represent probabilistic knowledge and relationships between random variables, and the decision network, which is an extension of the Bayesian network and can be used to make rational decisions based on uncertain data.

The system aims to teach a set of learning targets, which appear in the translation exercises. The learning targets may be individual words or larger constructions. The domain model represents their relations with the exercises.

The student model is the main contribution of this thesis. It is a Bayesian network that is partially structured after the domain model, as it also contains the learning targets and the exercises. It is, however, more complex, as it also stores probabilistic information about the user level, and

can be updated based on various kinds of evidence: the results of user's self-assessment and character recognition test, exercise outcomes and word look-ups in the dictionary.

The tutor model uses the information about a user's vocabulary knowledge in order to select the exercise that is most likely to lie in the student's zone of proximal development and maximise the learning outcome.

Chapter 4

Implementation

The presented system contains a hierarchically structured intelligent tutoring system with two layers: the client contains a simple dialogue system that presents exercises to the user and provides feedback, while the server performs cognitive modelling and selects exercises. The general design of the system and the way cognitive modelling is done has been presented in the previous chapter. The present chapter describes the structure of the system in more detail, and presents the actual implementation that was used to conduct the experiment presented in the next chapter. The implementation has an English name CHINESE IN CONTEXT, and a Chinese name 上下文中文 (*shàng-xià-wén zhōng-wén*, “context Chinese” when translated word-by-word, “up-down-language middle-language” when translated character-by-character).

4.1 Overview of the Implementation

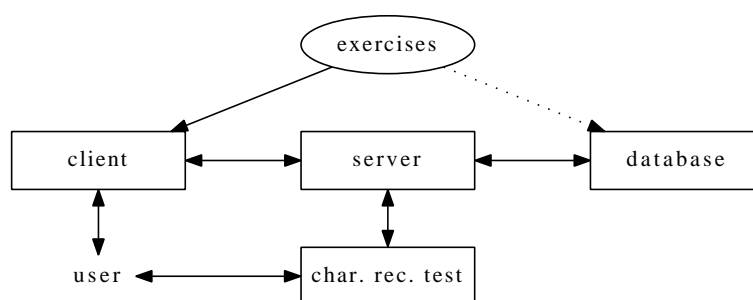


Figure 4.1: The structure of CHINESE IN CONTEXT tutoring system

Figure 4.1 shows the structure of the system. There are three main components: client, server and database. The character test is a supplementary component that works on the server side and interacts with the user before the main client is run. The system was written in Haskell (Peyton Jones et al. 2003), and the SQLite database¹ was accessed using SQL (ISO 1992). Since CHINESE IN CONTEXT is a web application, the

¹<http://www.sqlite.org>

client code had to be compiled into JavaScript, in order to be executed by web browsers. Therefore, the client was written in a subset of Haskell that is supported by Fay, a Haskell-to-JavaScript compiler.² The server uses Scotty, a Haskell web framework.³

The exercises are stored on the client side. However, as the dotted arrow indicates, the database was also based on the exercises: a Ruby⁴ script extracted dependencies between exercises and learning targets and used this information to create tables in the database and populate them with initial values.

The simple dialogue system that presents interactive exercises to the user is completely contained within the client, and is largely independent from other components – it can be used independently as an application run in a web browser.

The previous chapter provided information about the student modelling and the exercise selection. Those are actions performed by the server. Apart from the last section, this chapter concentrates mostly on the functions of the client. Apart from that, at the end of this chapter we will show how an SQL database provides a natural interface for creating a Bayesian network.

However, before we present the implementation of the client, we need to introduce some natural language processing concepts. This will be the topic of the next section.

4.2 Natural Language Processing

This section introduces natural language processing-related concepts that will be used in the description of the implementation of CHINESE IN CONTEXT: word segmentation and sentence distance metrics.

4.2.1 Word segmentation

In subsection 2.2.2 we have seen one of the major problems in processing Chinese: the word boundaries are not marked and may be ambiguous.

In general, we can divide word segmentation algorithms into two groups: **character-based approaches** and **word-based approaches**. Relative uniformity of morphology and syntax makes it possible to use the single-character based approach, which treats every character like a word. However, while this works fine for classical Chinese, for modern Chinese the precision is low (Wong, Li, et al. 2009, p. 44), and the word-based approaches are preferred.

There are three main sources of information for the word-based segmentation: dictionary, unsegmented corpora and segmented corpora. The **dictionary-based approach** relies on the list of known words to create a maximum match using a greedy search forward (Forward Maximum Matching) or backward (Backward Maximum Matching). FMM and

²<https://github.com/faylang/fay/wiki>

³<https://github.com/scotty-web/scotty>

⁴<http://www.ruby-lang.org>

BMM can be combined to improve accuracy (Wong, Li, et al. 2009). Corpora can give us information about which characters often appear together, and allow the use of **statistics-based approaches** (words are assumed to be character groups that are unlikely to appear by chance), the **comprehension-based approach** (bootstrapping with information about which morphemes are always free and which ones are always bound), and various kinds of supervised machine learning.

4.2.2 Sentence distance metrics

Let us assume that we have a sentence s (understood as a sequence of words $w_1 \dots w_n$) and a set of sentences V , and want to find a sentence belonging to V that is most similar to s . The degree of similarity of sentences can be defined in different ways. We will consider two possible metrics: edit distance and BLEU.

Edit distance between two strings of symbols s_1 and s_2 is the number of edit operations that need to be done in order to transform s_1 into s_2 . There are different versions of edit distance, depending on the definition of edit operation. In a common version, called **Levenshtein distance** after its inventor (Levenshtein 1966), there are three edit operations: removal, insertion and substitution of a single symbol. Edit distance has many applications in computer science, not only related to processing languages; in NLP it is probably most commonly used in spell-checking algorithms that look for candidate corrections of unknown words. In this case every character is understood to be a symbol. The edit distance can also be used on the word level, with every word being a symbol. A relatively straightforward algorithm for computing edit distance that uses dynamic programming has both space and time complexity $\Theta(mn)$ where m and n are the lengths of the respective sentences. There are more effective algorithms, but for the sentence lengths that occur in practice the difference is negligible.

Bilingual Evaluation Understudy (BLEU) is an algorithm closely related to NLP, and more specifically, to machine translation. Papineni et al. (2002) proposed it for computing a metric to evaluate machine translation, by comparing an automatically translated sentence against a set of reference translations made by humans. It is designed to work with multiple reference translations; however, for our purposes, we are interested in using it with one candidate sentence and one reference.

The algorithm is based on an insight that a good translation is more likely to share sequences of words with the reference translation than a bad one. Therefore, the basic idea is to generate all the n -grams of words in the candidate translation, and check how many of them appear in any of the reference translation. This ratio is called n -gram precision. If a particular n -gram appears more often in the candidate translation than in the reference, only the number of occurrences in the reference translation is taken into account. The core of the BLEU score is computed by taking a weighted geometric mean of the n -gram scores for different n . Papineni et al. (2002) note an important problem: very short candidate translations may get very high precision, yet if they are much shorter than the reference,

it is unlikely that they are good candidates. As a solution they propose a **brevity penalty**, which assigns lower scores to candidate translations that are shorter than the reference translation.

Following Koehn (2009) we will use a BLEU score with typical parameters: $n \in \{1, 2, 3, 4\}$, and uniform weights for the geometric mean. Such a score can be computed as follows:

$$\text{BLEU}(cand, ref) = \min \left(1, \frac{\text{length}(cand)}{\text{length}(ref)} \right) \prod_{i=1}^4 \text{precision}_i(cand, ref)$$

It is straightforward to use the BLEU score to find a sentence v from a set V that is most similar to the sentence s :

$$v_{sim} = \arg \max_{v \in V} \text{BLEU}(s, v)$$

In other words, for each v that belongs to V , we compute the BLEU score of s against v as the reference and choose the one that produces the highest score.

4.3 Character Recognition Test

The role of the character recognition test was to provide an objective measure of one's level. The test was built using a character frequency list in Modern Mandarin compiled by Da (2004). The characters have been divided into groups, depending on ranking in the frequency list: characters with ranks 1-190, 191-375, 376-750, 751-1500, 1501-3000 and 3001-6000 (with each interval being roughly twice as long as the previous one). 6000 was chosen as the upper limit, as this is an approximate number of characters that highly educated Chinese speakers know, as discussed in subsection 2.2.1. Apart from the frequency groups mentioned above, there was an additional group that provided distractors, and therefore it was made out of extremely rare characters that even native speakers are unlikely to know. The characters that have been chosen do not belong to the 20,992 most common ones that make up a block of so-called "CJK Unified Ideographs",⁵ established by the Unicode Consortium. The users have been presented 90 characters in random order, and asked to indicate as fast as possible whether they know a particular character. Each group was represented by a random sample of 13 or 12 characters.

The results of the test allow us to estimate the number of characters one can recognise. The percentage of recognised characters in each group indicates what percentage of characters in each group one can recognise. For example, if someone recognised 6 characters from the 1-190 group, out of 12, it would indicate that he probably knows about 50% of the characters in that group, that is, about 95 characters. Since the test subjects were extremely unlikely to really know the distractors, the number of distractors marked as recognised give an indication how likely someone was to mistakenly recognise

⁵<http://www.unicode.org/charts/PDF/U4E00.pdf>

an unknown character. This information was then used to appropriately decrease the estimate of known characters. For example, if 2 out of 12 distractors were marked as shown, we would conclude that about 2/12 (16%) of the non-distractor characters were also marked incorrectly, and decrease the final estimate by 16%.

The character test is presented at the beginning of the session. It is followed by a series of exercises, which are presented by the client. The user interface, design and implementation of the client is presented in the next sections.

4.4 Design and Implementation of the Client

As noted at the beginning of this chapter, each exercise is presented to the user by the client application that provides interactive feedback. The client runs in the user's browser; it can be used independently from other components. The only input that is required is the number of the exercise that is to be presented. At the end of the exercise, the client has an HTML form filled in with details of the session. It records time information about every character typed in by the user, every word that has been looked up in a dictionary, every system response, and the final status of the exercise (whether it was completed successfully or skipped). In CHINESE IN CONTEXT, the form with this information is submitted to the server, but the client may well be used as a component in a completely different application.

The structure of the client ITS is very simple: the domain knowledge is represented as the exercise list and there is no explicit student model. The student's actions are only recorded and passed to other components through the HTML form; they are not used directly by the client. The tutor model (with an implicit student model) provides interactive feedback to the student input.

4.4.1 User interface

Let us start the presentation of the client with its user interface. Figure 4.2 shows a screenshot of an ongoing session with the program and illustrates how an individual exercise is structured. The task is to translate the sentence *I work while having lunch*. Every word of the sentence is a hyperlink, which means that it can be clicked in order to look it up in a dictionary. The system uses the MDBG English-Chinese dictionary⁶ to show entries that contain a given word. The user has made three attempts at translating the sentence. As one can see from the figure, they used a construction that is correct only if there are two different subjects, and that cannot be used in this sentence. The relevant fragment was therefore highlighted in red. Then the user deleted the problematic fragment and again submitted the answer. The system found out that the user's answer does not contain a construction that is used to express that one thing is being done while another thing is

⁶<http://www.mdbg.net>

being done by the same subject. The system’s hint contained a hyperlink to a dictionary, which was clicked by the user, causing a dictionary entry to show up on the right-hand side. Then the user made a third attempt, using the construction from the dictionary. It is still, however, missing the expression for *eat lunch*, and therefore the exercise is not finished. A large number of correct answers is accepted: if possible, different orders of constituents are allowed, and synonyms are recognised. When the answer is correct, the system proceeds to the next exercise.

There is also an option of skipping the exercise. The user may have no idea about the correct answer to the exercise, despite the hints, or they may know that their answer is correct, despite the system saying otherwise. There are two *Skip* buttons to let the system designer know the reason for skipping.

4.4.2 Interactive feedback

The student-tutor interaction cannot be regarded as a fully fledged dialogue (as the student’s answers to the exercise are the only natural language input that is accepted by the system), but it can still be analysed within the framework of dialogue systems (Jokinen 2009, Jokinen & McTear 2009).

The error feedback is based on the principle of **reactive interaction**, that is, direct mapping a pattern in input into a reaction. Each exercise contains a set of strings that indicate an error if found in the user input. Matching such a string triggers a rule that gives the user a message describing what was written wrong and how it can be corrected. Since the rules are exercise-specific, this simple approach is expressive enough to describe many types of errors. Appendix A presents, among other things, error rules in the exercises that were created for evaluation of the system.

When no error rule matches, the system tries to provide feedback that leads the user towards the nearest answer. It is modelled after the **interaction modelling** approach to dialogue control. The dialogue context has structure: it is represented by a frame with a set of slots that are to be filled in in the course of the interaction. The next action is chosen from among the actions that lead to filling in the empty slots. In the tutorial dialogue, an answer to the exercise is a frame to be filled by the user.

The way of producing tutorial feedback is inspired by the constraint-based modelling (cf. subsection 2.4.2). There is a set of correct answers and three basic constraints: the user’s sentence needs to contain all the words that are in an answer, it cannot contain any superfluous characters, and the words need to be in the correct order. The dialogue manager chooses an answer that is closest to the user’s input, and checks these constraints. If they are met, the user’s answer is correct. If not, a constraint-specific rule is triggered, that leads the user one step towards the correct solution.

The user input needs to be split into words. For every exercise, there is only a limited vocabulary that the system needs to recognise and therefore a dictionary-based method is used. Forward maximum matching has turned out to work well in this application. The dialogue manager treats the sentence as a frame with slots to fill, so if there are words that are missing, it

Please translate the following sentence into Chinese:
[I work while having lunch](#)

- **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).

我一边吃饭，一边工作

Submit

Skip (I give up) >>

Skip (I'm sure my translation is correct) >>

Your input: 吃饭的时候 我工作

时的时候 pattern can only be used if there are two different subjects. There is only one subject here ('I'), so you need to use another construction.

Your input: 吃饭 我工作

You're missing a word/phrase: [on the one hand, on the other hand, doing while](#).

Your input: 我一边吃饭，一边工作

You're missing a word/phrase: [eat lunch](#).

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.

Number of remaining sentences before the final test: 14

- ☐ 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:

Go

一边 yī biān

one side / either side / on the one hand / on the other hand / doing while

Kinesiskurs i Stavanger.

jiaoxie.no

Kveidskurs og bedriftskurs - Språk og kulturkompetanse

Dictionary · [Google Translate](#) · [Help](#)

By [MDBG](#) 2014

Figure 4.2: Ongoing session with the program

chooses the first (leftmost) of them, and informs the user that it is missing. The language generator does not provide the Chinese word, but only an English version, and a link to a dictionary that the user may use. If there are superfluous characters, a rule is triggered that asks the user to remove them one by one. Finally, if the two previous constraints are met, but the word order is incorrect, the third rule invokes a message that explains the correct word order (again, without giving specific Chinese words) and asks the user to change it.

4.4.3 Representation of the answer set

The set of correct answers to the exercises is stored in the form of a formal grammar. They are stored in a format that is equivalent to the Backus–Naur Form and is able to express context-free grammars. Answers to the exercises that were created to conduct the experiment presented in the next chapter did not take full advantage of this possibility: the rules had a simple two-level structure, with a rule producing the sentence structure at the first level (partly filled with some terminals), and rules producing other terminals at the second level. Regular grammars are not expressive enough to describe natural languages (Chomsky 1956). In this case, however, this is not a problem, because we are not describing all potential sentences in a language, but only a set of answers to an exercise. We can probably assume that this set is always finite, and in such a case a regular grammar is expressive enough. The regular grammars are, however, not practical – they do not allow more than one non-terminal on the right-hand side (Jurafsky & Martin 2008) and therefore require using rules that are not natural from the linguistic perspective. The choice of a context-free grammar to describe the answers allows using rules that are natural and much more concise. The actual rules that are used to describe the answers are presented in the appendices.

4.4.4 Calculating the sentence distance

The dialogue manager leads the user towards the answer that is closest to the input. The two distance metrics described in subsection 4.2.2 were considered. During preliminary testing, the Levenshtein distance was used. However, the BLEU algorithm turned out to work better with problems with the word order and was therefore used in the final system.

The difference between the two distance metrics can be seen in the following example. Let us take an exercise that asks to translate the sentence “I am going to Beijing tomorrow”. Since 明天 *míngtiān* and 明儿 *míngr* both mean “tomorrow” and can be used in this context, both of the sentences below can be considered correct answers:

我	明天	去	北京
<i>wǒ</i>	<i>míngtiān</i>	<i>qù</i>	<i>Běijīng</i>
我	明儿	去	北京
<i>wǒ</i>	<i>míngr</i>	<i>qù</i>	<i>Běijīng</i>
I	tomorrow	go	Beijing

“I am going to Beijing tomorrow”

Let us suppose that a student wrote all the words correctly, but in a wrong order, e.g.:

*我	去	北京	明天
<i>wǒ</i>	<i>qù</i>	<i>Běijīng</i>	<i>míngtiān</i>
I	go	Beijing	tomorrow

If we compute the word-level Levenshtein distance between this answer and the two correct ones, we will find out that the distance to both correct answers is the same. Changing the student’s answer to a correct one requires removing 明天 from the end of the sentence, and inserting either 明天 *míngtiān* or 明儿 *míngér* in the second position, therefore the distance is 2 if we compute it on the word level, and 4 if we compute it on the character level. Intuitively, we want the correct answer which uses 明天 *míngtiān* to be closer – since it is a correct word for “tomorrow”, it would be confusing if the system asked the user to change it to a synonym, 明儿 *míngér*.

BLEU handles these situations better, as its computation of similarity is based on the number of common n -grams. If we compute it at the character level, we will see that most n -grams in the student’s sentence are present in both correct answers. There are, however, two more n -grams that are only present in the answer that we want to be closer, namely 天 *tiān* and 明天 *míngtiān*, and they contribute to the higher similarity score for this sentence.

4.5 Relational Representation of the Bayesian Network

In the previous chapter the approach to the student modelling was described; the core of the student model consists of a Bayesian network. As mentioned in section 4.1, the network is stored in CHINESE IN CONTEXT in a SQL database. The use of a relational database is not just an implementation detail, the connection between Bayesian networks and relational databases is deeper and has theoretical background.

Wong, Xiang & Nie (1995, p. 1) point out important equivalences: “Relational databases manipulate tables of tuples, and Bayesian networks manipulate tables of probabilities. Relational databases answer queries that involve attributes in multiple relations by joining the relations and then projecting to the set of target attributes. In Bayesian networks, joint distributions are defined by products of local distributions, and belief updating computes the marginalization of joint distributions.” The details of these equivalences are beyond the scope of this thesis. The practical advantage of using an SQL database to represent Bayesian network can be seen in the following example.

Assume that the exercise **ex88** has been finished successfully and the learning target **cons65**, which appears in this exercise, has not been looked up by the user. The new probabilities of user knowing the learning target

cons65 can be computed as a matter of a single SQL statement (the results are not normalised, so the each value needs to be divided by their total sum in order to obtain probabilities):

```
SELECT cons65, SUM(cons65_p*ex88_p*lookup_c65_e88_p) AS p
FROM cons65 NATURAL JOIN ex88 NATURAL JOIN lookup_c65_e88
WHERE ex88='Y' AND lookup_c65_e88='N' GROUP BY cons65;
```

As we have seen, calculating a new updated probability requires many steps. In the case of more complicated structures, the order of execution also plays a role. Thanks to the equivalence between the structure of the relational database and the Bayesian network, we can use SQL to formulate the inference in a high-level, declarative fashion. This query belongs to a class called **Marginalise a Product Function** (MPF) queries. Bravo & Ramakrishnan (2007) have shown that the query optimiser can use efficient exact inference algorithms, such as variable elimination, to answer such queries.

Chapter 5

Experiment

The goal of the experiment was to investigate the empirical effects of using the cognitive model to select the next exercise. Before the experiment could be conducted, the system was tested on users, to make sure that subsequent users will not have problems with using the system without assistance. In the data collection phase, a number of people were recruited to try a simplified version of the system that did not do any cognitive modelling, and chose exercises randomly. The collected data allowed the initialisation of parameters of the cognitive model. The actual experiment compared this model with a baseline, which chose exercises without cognitive modelling.

The first section of this chapter presents the process of testing and data collection. In the second section, we will look into how parameters of the model were estimated. Subsequent sections present the experiment that was performed and discuss the results.

5.1 Testing and Data Collection

The first version of the system consisted of the following elements:

- an initial questionnaire,
- a character recognition test,
- the main program,
- a final questionnaire.

The initial questionnaire collected information about users' self-assessment of their written and spoken proficiency in Chinese, e-mail address, gender, the amount of time they have been learning Chinese, as well as the standard of Chinese characters they were most familiar with (simplified or traditional).

At the first stage, 3 second language learners of Chinese took 8 exercises, under our supervision. The observation showed that the error anticipation using pattern matching worked as expected – when an erroneous answer was provided, the system provided useful feedback about what should be corrected. The overall experience with the user interface was positive. The

users had several issues, e.g. highlighting every individual word confused one of the users, and he did not see at first glance that it was the sentence that was to be translated. The issues were addressed by more detailed written explanations that made sure that users knew what to do. Subsequent user sessions were conducted online, without our presence, therefore those first sessions were important to make sure that the user interface is self-explanatory.

In the next stage, 92 exercises were created, in addition to the previous 8. Subsequently, 3 native Chinese speakers were consulted on construction of the exercises, and provided more ways of expressing a given meaning, expanded lists of synonyms, and verified that error rules indeed catch incorrect ways of constructing a sentence in a given context.

60 second language learners of Chinese were then asked to fill in the questionnaire, take the character recognition test, use the program on randomly chosen exercises, and fill in a questionnaire. In the final questionnaire the users were asked if the session was interesting and if they wanted to repeat it in the future. The data collected in this phase provided initial values for some of the parameters of the learner's cognitive model.

54 of the users filled in the initial questionnaire and did the character recognition test. They were then asked to do at least 20 exercises, but they were allowed to do more, up to 100. There were 5 users who filled in the questionnaire and did the test more than once. The subsequent questionnaire and test results were discarded. 4 users did not continue to the exercises because of a technical issue – they were accessing the application from mobile phones, which are not supported. There were 11 users in total who did not finish a single exercise, aside from those who used mobile phones, it is not clear whether it was due to technical issues, or some other reasons. Among the 43 users who started the exercises, the majority (31) did at least 20 of them, as they were requested.

5.2 Parameter Estimation

5.2.1 User levels

One of the aims of the data analysis was to find a single variable that could represent language level of the users in the context of exercises, that is, aspects of the level that are relevant when the user does the exercises and that can help to find an exercise that is not too easy nor too difficult. The starting point for the variable indicating user level, are CEFR levels (Council of Europe 2001) that were also used for the self-assessment: A1 Newbie, A2 Elementary, B1 Intermediate, B2 Upper-intermediate, C1 Advanced, C2 Proficient.

Figures 5.1, 5.2, 5.3 and 5.4 present the data with regard to different variables. There are 2 categorical variables: users' self-assessment of their written proficiency, and self-assessment of their spoken proficiency. In addition, we extracted 3 continuous variables: estimated number of characters that the learner knows, skip ratio (number of exercises skipped divided by the total number of exercises the user went through) and look-up

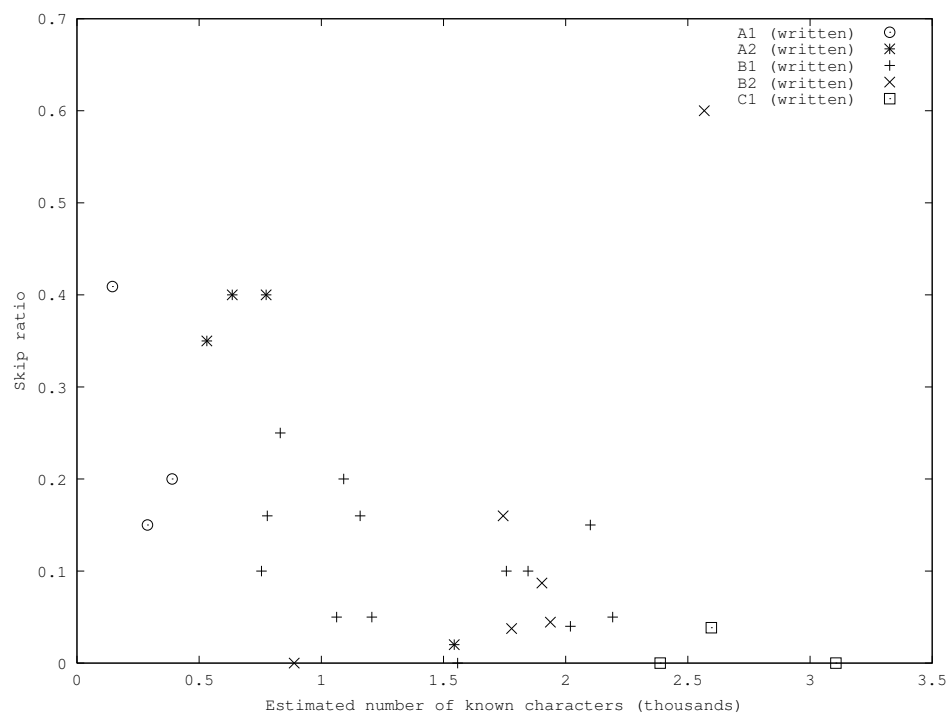


Figure 5.1: Skip ratio, estimated number of known characters and self-assessment of written proficiency

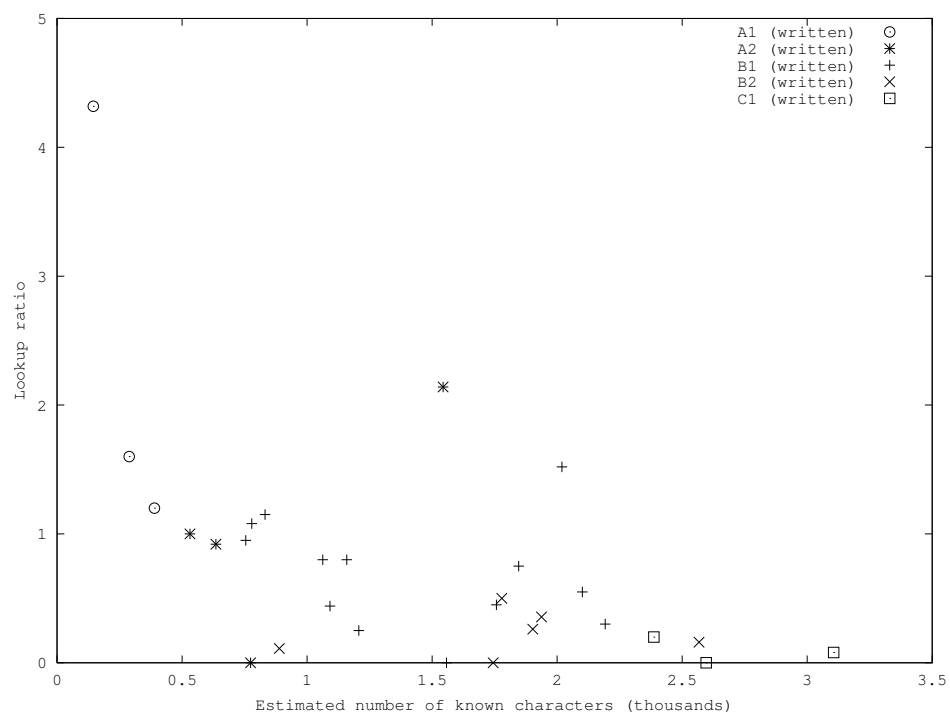


Figure 5.2: Look-up ratio, estimated number of known characters and self-assessment of written proficiency

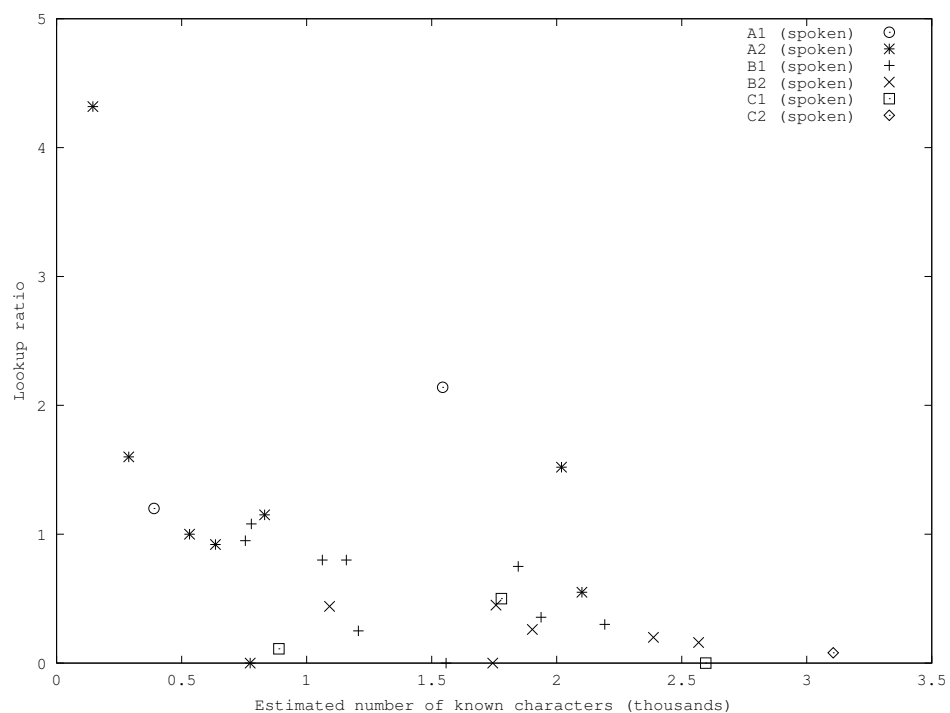


Figure 5.3: Look-up ratio, estimated number of known characters and self-assessment of spoken proficiency

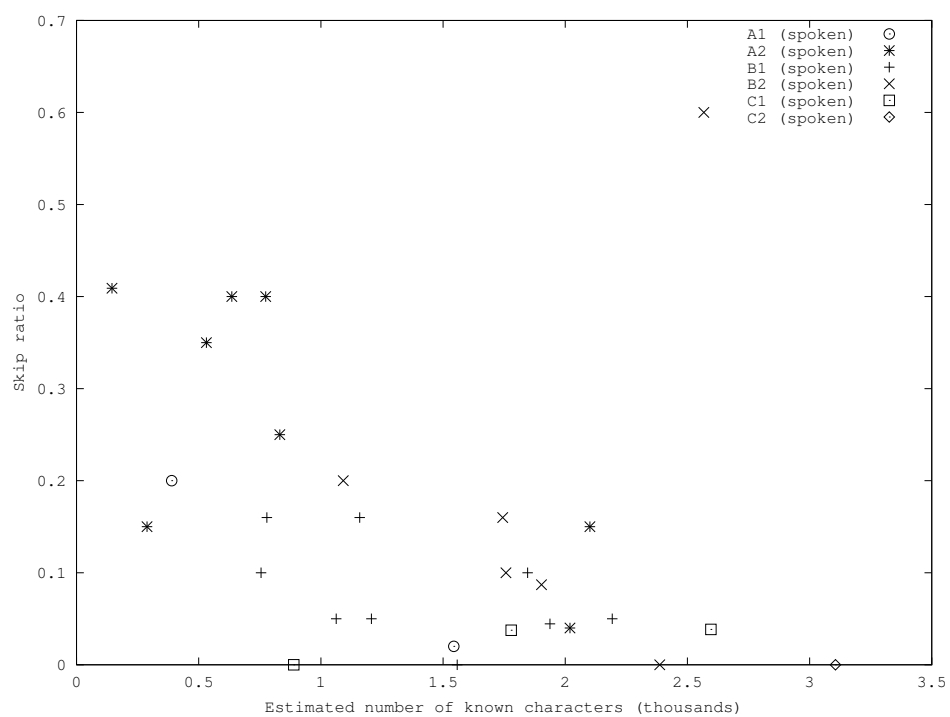


Figure 5.4: Skip ratio, estimated number of known characters and self-assessment of spoken proficiency

ratio (number of exercises in which some word has been looked up divided by the total number of exercises the user went through). If we consider grouping according to user’s self-assessment, we can measure intra-group variance according to different dimensions. It turned out that the dimension with highest intra-group variance was the estimated number of characters. The figures show this dimension combined with each of two other continuous dimensions – skip ratio and look-up ratio.

We can see that levels A, B and C are quite separated in the *number of characters* dimension, especially for the written levels, and there are also differences in other dimensions. We also see that 3 users who assessed their level as C1 or C2 did very few look-ups and skipped few exercises. Those who assessed their level as A1 or A2 know below 1000 characters and have high look-up ratio, and those who assessed their level as B1 or B2 are in between the two former groups with regard to those parameters. It is hard, however, to find a pattern for sub-levels. The data does not show any clear patterns with regard to sub-levels 1 and 2 of any of the levels A, B or C. Clusters A and C are quite homogeneous, apart from individual outliers, while cluster B seems to have two sub-clusters along the *number of characters* axis – one concentrated around 1000 characters, and another concentrated around 2000 characters. Therefore, even though division into these clusters does not match the division between those who assessed their level as B1 and those who assessed their level as B2, we decided to differentiate between these two clusters, and call the one concentrated around 1000 characters B1, and the other one B2. This is based on an assumption that while most users can decide whether their level is beginner, intermediate or advanced, differentiation between sub-levels such as intermediate and upper-intermediate is subject to much subjective interpretation, and results of a character recognition test are more likely to give some insight here. Neither cluster A nor cluster C had such clear sub-clusters. Thus, the final list of clusters that were used for initialisation of the parameters of the exercise was as follows: A, B1, B2, C.

5.2.2 Conditional probabilities

The estimation of conditional probabilities was based on the division of users into four level groups described in the previous section. Under the assumption that the character recognition test results are normally distributed within each group, the results were used to estimate the distribution of the number of known characters given the user’s actual level.

Information about look-ups was used to divide learning targets into difficulty classes, named after user levels. For every level X, class X contains learning targets that were looked up by some users whose level is X, but not by those at higher levels. The targets that were never looked up by anyone were assigned to the class A0. The members of the A0 class were removed from the learning targets list, as they most likely contained items that had not caused problems even to beginners.

Since look-ups are only an indirect evidence of not knowing a word, it was not possible to use them to directly estimate actual probabilities of

LT level	User level	P	LT level	User level	P
A	A	0.3	B2	A	0.1
	B1	0.9		B1	0.2
	B2	0.99		B2	0.3
	C	0.99		C	0.9
B1	A	0.1	C	A	0.1
	B1	0.3		B1	0.2
	B2	0.9		B2	0.3
	C	0.99		C	0.4

Table 5.1: Conditional probability tables for learning targets (LT) at different levels

knowing words without making additional assumptions. Therefore, we took a simpler approach, and manually filled in the conditional probability tables, separately for each difficulty class.

Table 5.1 shows conditional probability tables for the four difficulty classes. For example, a learning target that was estimated to have difficulty B2 used a CPT in the upper right corner. As we can see, the system assumed that users whose level was estimated at level B2 were assumed to know this word with probability 0.3. The same probability was used in the cases where the LT level was the same as the user level, except for level C. In this case the value was made slightly higher, because the interaction analysis suggested that it was more likely that those users actually knew the words they looked up, and looking up was used more to test the system rather than to actually find unknown words. As expected, the probability of knowing words increases as the user level goes up, and decreases as it goes down. We should also note the asymmetry: the probability that an easy learning target (level A) is known by an advanced user (level C) is very high (0.99), which means that probability that it is unknown is only 0.01. On the other hand, the probability that a difficult learning target (level C) is known by a beginner (level A) is 0.1, which is 10 times more than 0.01. This asymmetry can be explained as follows: If we know that somebody does not know a very frequent word, it is strong evidence against the hypothesis that this person is an advanced learner, but knowing an infrequent word is not such strong evidence for being at an advanced level. If we take learning English as an example, it is hard to imagine an advanced learner who does not know the word *house*. On the other hand, there is nothing particularly unusual with a beginner who knows a relatively infrequent word, say *ophthalmologist* (especially if they live in an English-speaking country and wear glasses).

The above tables were filled in after looking at look-up patterns of users at different levels, but the actual values are just guesses. It must be stressed, however, that they are simply prior probability values, and they are updated after each exercise. That means that as a learner does more and more exercises, the values become more adapted to a particular user’s vocabulary knowledge, and their initial values become less important. In other words,

Level	0	1	2	3	4	5
A	0	16	8	4	2	1
B1	0	4	8	4	2	1
B2	0	1	2	4	8	16
C	0	1	2	4	8	16

Table 5.2: Utility values, based on the user’s level (vertical dimension) and the number of unknown learning targets in a sentence (horizontal dimension)

we use a Bayesian interpretation of probability: it is important how updating is done, not the way the initial values are chosen (though less updating is needed when these values are reasonable).

The data also showed a very high skip ratio (over 0.5) for 6 exercises. This may indicate that these exercises accept too small a number of possible alternative translations. The best solution would be to go through user interactions with these exercises and analyse the feedback the system provided. However, due to lack of sufficient time, these 6 exercises were excluded from the actual experiment. 94 exercises were left.

5.2.3 Utility values

As described in chapter 3, our approach assumes that learning targets may be either known or unknown, and probability expresses our uncertain knowledge about their actual status. Under such assumptions, an exercise is definitely too easy if the learner knows all the learning targets that appear in it. We cannot expect the learner to have any problems with translating such a sentence into Chinese, but it still will not lead to learning – if everything is already known, nothing can be learnt. It is harder to define when an exercise is too difficult. If a sentence has many unknown constructions, it is unlikely that they can all be learnt at once. It is, however, uncertain where the borderline is. It is also something that cannot be inferred directly from the data, therefore we decided to hard-code the utility of exercises, depending on the number of unknown words.

As we saw in subsection 2.1.1, each learner has their own zone of proximal development, and learning can be seen as progression from being other-regulated to being self-regulated. More advanced learners have more experience with becoming self-regulated, and therefore we can assume that their ZPD is generally larger. In other words, more advanced users are more likely to remember many unknown words after looking them up than beginners. That is why the utility function depends on the level of the user. Table 5.2 shows actual values of the utility function used by the system.

5.3 User Evaluation

5.3.1 Experimental setup

In the experiment, a set of 94 exercises were used. They contained 91 different constructions, selected as learning targets. The point of the system was to use the cognitive model to make a choice between repeating some learning targets, and not repeating others (when it became likely enough that they have already been learnt), therefore some constructions were repeated among different exercises, with about 3 exercises per learning target on average. Very frequent and basic words, such as pronouns, and words that were not repeated among exercises were not considered learning targets. The program is aimed at people who can already construct simple sentences, and the pronoun system in Chinese is simple, so this is not problematic for the assumption that it is the learning targets that determine the difficulty of the sentence.

The participants were recruited by announcements that were made at Chinese language classes at the University of Oslo, at forums devoted to learning Chinese and general language learning, and on Facebook. Second language learners of Chinese are relatively scattered, and there is no easy way to get direct access to hidden populations of such users. Therefore, to recruit enough participants, snowball sampling was used too – the users were encouraged to share invitations to use the system with their friends on Facebook, Twitter and Google+.

The sessions with the participants had the following sequence:

- **Assignment.** The participants were randomly assigned to the actual system or to the baseline (but were not informed about it in any way).
- **Initial questionnaire.** The participants filled in a questionnaire, which asked about self-assessment of their written and spoken proficiency in Chinese, e-mail address, gender, the amount of time they have been learning Chinese, as well as the standard of Chinese characters they were most familiar with (simplified or traditional).
- **Character recognition test.** The users were presented 90 randomly chosen characters from different frequency ranks and were to tick check-boxes near the characters that they knew.
- **Main session with the system.** The participants were presented 14 exercises. The choice of the next exercise was the only difference between the system and the baseline. The system used a decision network to select the exercise with the highest utility, while the baseline chose a random unseen exercise at the user's declared level.
- **Post-test.** The participants were asked to translate a stratified random sample of the learning targets, with 6 random items from each of the 3 strata: A, B1, B2+C. There were only 3 learning targets in class C, therefore classes B2 and C were joined.

- **Final questionnaire with subjective assessment.** The users were also asked to fill in a final questionnaire, where they were asked if the exercises were suited to their level and how much they had learnt during the session.

5.3.2 Baseline

A baseline was needed to compare the presented solution to the exercise selection problem with a simple, yet reasonable way of choosing the next exercise. The baseline that was created for the purpose of the experiment did not use the cognitive model, but had a pre-defined notion of level of the exercises. For each exercise, we took all its learning targets and found the one with the highest level. It can be assumed to be the most difficult part of the sentence, and therefore determines the level of the exercise. After assigning a level to each exercise, the baseline employed the user's declared written level, and chose a random unseen exercise at that level. If there were no unseen exercises at that level, the baseline was to take exercises at higher levels, and if all higher levels were exhausted – also lower ones. In practice, during the experiment there were enough exercises at the declared user level apart from the users who stated their level as C1 or C2. There were only 3 exercises at level C, and therefore subsequent exercises were chosen from level B2 instead.

5.3.3 Results

60 people filled in the initial questionnaire; 48 of them did the character recognition test and did at least one exercise. 33 of them went through all the assigned exercises and filled in the post-test and the final questionnaire. After discarding those who had used the system before (by comparing emails with those which were given during previous tests of the system), and two participants who had Chinese as their mother tongue, 24 participants were left.

The results are presented in Table 5.3. There were 15 participants who were randomly assigned to the system, and 9 who were randomly assigned to the baseline.

The post-test results were compared according to estimated user's level. Both the users of the system and the users of the baseline were grouped by their estimated level, as indicated by the Bayesian network.

The difficulty assessment was performed by asking the question: "The application tried to adapt to your level, and give you sentences that are not too easy, nor too difficult. How did it work for you?". Those who chose the answer that the level was "OK" were assigned value 0, those who chose the answers that it was "a bit too easy" or "a bit too difficult" were assigned value 1, and those who chose the answers that it was "way too easy" or "way too difficult" were assigned value 2.

The second subjective assessment question was stated as follows: "How many words/constructions do you think you have learnt in this session?".

	User level	System (15 users)		Baseline (9 users)		p-value
		Mean	Std. dev.	Mean	Std. dev.	
Post-test results (the 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; the lower the value, the better the program was aligned with users' expectations)		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 (the fraction of users who did not finish doing the exercises)		0.38	0.49	0.31	0.46	0.6451

Table 5.3: Results of the experiment

The four possible answers were assigned numerical values: “none at all” (0), “just a few” (1), “some” (2) and “a lot” (3).

The drop-out rate was calculated by taking those who started doing the exercises, but did not finish them, divided by the total number of users who started doing the exercises.

A two-sided unpaired two-sample t-test was performed to see whether any of the results are statistically significant. No statistically significant differences have been found in the post-test for users at any level, drop-out rate, nor were the differences to answers to the first question. There is, however, a statistically significant difference ($p < 0.05$) between the users of the system and the users of the baseline in answers to the second question.

Apart from the qualitative evaluation presented above, the users were presented with the possibility of giving free-form feedback. Apart from that, some of the users contacted the author directly. There were several types of feedback. Some users simply stated that they liked the application. Others pointed out particular synonyms or phrases that should have been recognised, or simply stated that the program should accept more variants. The third group raised other issues they had with the program. The suggestions included displaying a progress bar that would visualise how much work is left, showing a possible correct answer when an exercise is skipped, and there were complaints that the built-in dictionary did not provide all of the required words. Feedback from some users indicated that they perceived it primarily as a grammar tutor, and not a vocabulary tutor. The feedback suggested that most of the users understood the purpose of the program as a learning tool, but two relatively advanced users (at B2 and C levels) suggested that it would be more suitable for testing.

5.4 Discussion

The evaluation of objective effects (knowledge of learning targets after using the program) did not show statistically significant differences. It might have been caused by the fact, that this type of evaluation had to be done separately for users at different levels, as only within levels is it possible to assume that user’s prior knowledge is similar. Among the 34 users who used the program and took the post-test, 10 were at level A, 8 were at level B1, 4 were at level B2 and 3 were at level C. The subjective effects, on the other hand, could be compared for the whole populations that filled in the final questionnaire. It is therefore possible that the lack of difference in objective measures was due to too small sample sizes. An experiment with a larger number of participants is needed to investigate this question. Such an experiment may also provide enough data to tune the parameters of the hidden nodes of the Bayesian network, that were set manually in this experiment.

There are reasons to believe that the baseline chose exercises that were generally easier than the ones chosen by the system. For the baseline the difficulty of a sentence was equal to the difficulty of its most difficult element. The system chose exercises in a more informed way, based on what it knew

about a user's knowledge. From the data available to the system it was easier to deduce which words were known (successful exercises provided strong evidence) than to deduce which ones were not known (the evidence provided by skips and look-ups was relatively weaker). Therefore, the system had more information about words known by the user than could be deduced from the user's level alone. It could use this information to select some of exercises that would otherwise be regarded as too difficult.

Since the exercises chosen by the baseline were probably easier, we can assume that the baseline played its role well to provide a lower bound for difficulty. We cannot be sure, however, that it provided an upper bound. The aim of the system was to intelligently choose exercises at the right level. The experiment has shown that the system that intelligently chooses exercises performed not worse, and probably slightly better, than a baseline that consistently chooses easy exercises. Further experiments should compare the system against another baseline, that consistently chooses difficult exercises. For example, it could choose exercises that are one level higher than the level of the exercises chosen by the original baseline. If the system also performs well against such a baseline, it will provide much stronger evidence in favour of our approach.

The users were randomly assigned to the system or to the baseline at the beginning of each session. However, due to the relatively low number of users, there was quite a big disproportion between the two groups: 15 users were assigned to the system, and only 9 to the baseline. This made it impossible to compare the objective post-test results for the whole population. For example, all 3 C-level users happened to be assigned to the system. As we can see in Table 5.3, they all had the highest possible result in the post-test. If we compared the post-test results between the system and the baseline for whole populations, the calculated results of those who used the system might have been disproportionally high because of this coincidence. This problem is much less likely to occur with more participants. Another solution is to divide users into levels first, and randomly assign half of each level to the system, and the other half to the baseline.

The results obtained by the users at advanced and upper-intermediate levels (groups C and B2) suggest that even the most difficult exercises provided by the program were easy for users at those levels. It may be a problem with this particular set of exercises, but it may also mean that the tutoring model used in the program is not suitable for more advanced users. A set of more challenging exercises needs to be created to investigate this. Creating them would be more difficult, as they would have to accept a larger set of answers that more advanced users are likely to provide. Some of the users suggested that the program would be better for testing, rather than for learning. This may mean that this teaching method works only for beginner and intermediate-level students. On the other hand, there were many users at all levels who indicated that they liked the program and that they had learnt a lot. It is therefore likely that the suitability of the tutoring model has less to do with the actual level, and more with one's individual learning style. An experiment with more challenging exercises on upper-

intermediate and advanced users would give us more data and help to find out which of the possibilities is more likely.

5.5 Summary

In this chapter we presented an experiment that investigated measurable effects of integrating a learner model into a tutoring system. Before conducting the experiment, testing and data collection was performed. The collected data were used to initialise parameters of the model. In the experiment, the system with a cognitive model was compared with a baseline that chose exercises according to user's declared level, without using cognitive modelling. The parameters that were compared belong to two groups: objective assessment of users' knowledge after using the system or the baseline, and users' subjective self-evaluation of the difficulty of the exercises and the learning outcome. The system was never significantly worse than the baseline. In most cases, the differences were not statistically significant; however, the users' subjective assessment of how much they learnt showed significantly better results for the users of the system with a cognitive model.

The results allow us to conclude that the system performed at least comparably to the baseline. The most significant problem with the presented experiment was the small number of participants.

The baseline provided a good lower bound for the difficulty of the exercises; however, another baseline providing consistently more difficult exercises would allow better assessment of the system's performance. Conversely, with the small number of participants, the number of experimental groups had to be restricted, and therefore only one baseline was used.

The lack of statistically significant difference for the objective outcomes at different levels is likely to be caused by a small number of participants, too. Since the users' self-assessment of the learning outcome is higher for the system than for the baseline, it is likely that with enough participants this difference should be measurable.

We can see that the results of the experiment indicate a direction for further investigations. A very important factor is the number of participants. Larger groups can allow for more sophisticated experimental design, possibly with more than one baseline, and can lead to a more definite conclusion about the objective empirical results of the system.

Chapter 6

Conclusion

6.1 Summary of the Thesis

This thesis investigated the following questions, as listed in the introduction:

1. What can we learn from theories of learning and second language acquisition for the development of language tutors?
2. How to structure the student model of a language tutor as a Bayesian network representing the learner's vocabulary knowledge, and use this model to select the most appropriate exercises?
3. Does using this student model lead to measurable improvements in the learning outcomes of this tutoring system?

The first question was addressed in chapters 2 and 3. Chapter 2 introduced some important concepts in several fields, which included learning theories, second language acquisition and intelligent tutoring systems. These concepts were then used in the first section of chapter 3 to motivate the general design of the system.

Subsequent sections of chapter 3 were related to the second question. They introduced probabilistic graphical models and showed how they can be used to create a student model and a tutor model, and how the latter can use the information stored in the former.

Chapter 4 presented an implementation of the system, which was then used to conduct an experiment, and allowed evaluation of the empirical effects of using the system. The experiment and its results are presented in chapter 5, which addresses the third question.

The pushed output hypothesis served as a theoretical background for the exercise design. The translation exercises designed and created as a part of this thesis are based on this hypothesis; they push the learner to formulate sentences, and provide dynamic corrective feedback. The use of this type of feedback is motivated by studies that suggest its effectiveness. In the exercises, the feedback leads the learner towards the closest correct answer: the system first makes sure that the user has written all the words that need to appear in the sentence, and then makes the user group them into

larger phrases, and finally put them in the correct order. Such a bottom-up approach is motivated by the lexical approach to language teaching: according to this approach, sentences are usually constructed by combining chunks of words that the learner already knows, rather than by filling in sentence patterns with individual words, as suggested by the traditional grammar-based approaches. This thesis provided a new, theoretically motivated design for implementing such types of tutoring systems.

The most important contribution of this thesis is designing, creating and evaluating a student model. Learning is more effective when exercises are adapted to the user's level. The approach to the student modelling in this thesis is based on learning theories, most importantly, on Vygotsky's concept of the zone of proximal development (ZPD), introduced in chapter 2: the range of activities that the learner can do with assistance, but not alone. An effective tutoring system should provide exercises that lie within the ZPD: the ones that are neither too easy nor too difficult.

In the case of a vocabulary tutor, the student's knowledge of various words and constructions are important factors that influence that student's ZPD. Since the vocabulary knowledge is not directly observable, we employed probabilistic techniques, introduced in chapter 3, namely Bayesian networks and decision networks. These techniques allow us to elegantly represent the available uncertain information about the user's knowledge, and use this information to select exercises.

The system has been implemented and evaluated in a user experiment. The experiment described in chapter 5 compared the system with a baseline that chose random exercises at the user's stated level without any cognitive modelling. The empirical performance of the system was comparable to the baseline in most aspects. The lack of significant differences between the system and the baseline in these aspects might have been caused by the relatively small number of participants in the experiment. Nonetheless, the users' self-evaluation of how much they have learnt showed a statistically significant difference between the system and the baseline, which suggests that significant differences in other aspects may possibly be detected in experiments with a larger number of participants.

6.2 Future Work

There are several directions in which the work done in this thesis can be extended.

6.2.1 Domain modelling

The thesis focused on the student model and its use to select the next exercise. The domain model, presented in subsection 3.3.1, was quite simple: it only modelled relations between the learning targets and exercises. However, the formalism of semantic networks is flexible, and allows the domain model to be extended in several ways.

Even though we chose a subset of constructions appearing in the exercises as the learning targets, it may be beneficial for the system to take into

account all the constructions that are present, not only the ones that the system actively tries to teach. This can be done in a straightforward manner, as it only requires the introduction of new nodes (and associated edges) for these constructions.

In the domain model, no relations between the constructions were modelled; they were assumed to be independent. Our choice of the learning targets tried to reflect this assumption: we chose constructions that did not have any obvious relation with each other. A system with a larger number of exercises will necessarily teach items that are somehow related. The system teaches constructions, which may contain more than one word. The individual words that appear in the constructions also need to be taught. The domain model should reflect relations between constructions and individual words (or smaller constructions) that appear in it, as this has an influence on how and when they should be taught. Also semantic relations, such as synonymy or antonymy should be modelled: they are psycholinguistically motivated (Beckwith et al. 1991), and are therefore likely to have an influence on how words and constructions are learned and retained.

Another type of information that may be useful to model is frequency. Frequent constructions are easier to learn. It may, therefore, be advantageous to divide constructions into several classes, based on their frequency. These changes can then be reflected in the student model.

There are also possibilities to better adapt the domain model to the features of Chinese. In this language, words consist of characters that are much harder to remember than, for example, letters in alphabetic scripts. Knowledge of characters therefore has a strong influence on learning words: words with unknown characters are harder to learn than words that consist of characters that the learner already knows. If the domain model stored the relationships between words and characters, it would be possible to track such influences. Moreover, just as in the case of words, it may be useful to take advantage of character frequency information.

All these changes to the domain model can be relatively easily reflected in the student model. Part of the Bayesian network in the student model is directly structured after the domain model, so it will have to be changed accordingly.

6.2.2 Memory modelling

The current system is concerned with modelling user's knowledge during one session. As we mentioned in subsection 2.4.2, a useful tutoring system should ideally model changes to the user's knowledge over extended periods of time. It needs to be investigated how to integrate long-term memory effects into the cognitive model. Evaluation of such a model requires conducting experiments over longer periods of time, which was not possible to do within the scope of a Master's thesis.

The long-term perspective requires the drawing of an explicit distinction between long-term and short-term memory. This distinction may be introduced into the student model by creating two nodes for each learning

target. One of them can represent whether a construction is in the student's short-term memory, and the other can represent its presence in the long-term memory. We want to express various relationships, for example, if an unknown construction was looked up and used to answer an exercise, it is probably in the student's short-term memory, but not in the long-term memory. As the construction is used again, the probability of committing it into the long-term memory increases. Expressing such relationships may be difficult in a standard Bayesian network, so this approach may require using its extension that allows it to explicitly model time: a dynamic Bayesian network (see e.g. Russell & Norvig 2010, ch. 15).

However, such a model might not be sufficient. We saw in subsection 2.4.2 that effective learning requires repetition, most likely in increasing intervals. We cannot assume that the constructions that have been learnt, even if they are in the long-term memory, will stay there infinitely. Therefore, it would be very beneficial to have an explicit model of forgetting.

At the end of subsection 2.4.2 we mentioned an extension to ACT-R developed by Pavlik (2006). It assumes that each practised item gets an increased level of activation, and provides a function to compute how this activation decreases in time (Pavlik & Anderson 2010). In this way the model retains the general spacing principle from the approaches discussed at the end of subsection 2.4.2. Extending the student model with a model of forgetting based on this approach is certainly an interesting research direction.

6.2.3 Exercise encoding

Currently each exercise consists of one English sentence, and one or more patterns that can be used to generate possible answers. The representation of the possible translations is quite compact. However, each new English sentence requires at least one new pattern, even if its structure is shared with preexisting ones. With the 100 exercises that were created as a part of the implementation, this was not a significant concern. However, if more exercises are to be created, a way of sharing the structure should be developed. It would allow for the creating a large number of exercises in a short time. For example, there are many ways of modifying the sentence *He sold his piano yesterday* without changing the general pattern: *He sold his bike yesterday*, *He sold his piano on Monday*, *He bought his piano yesterday*, and many more. With a way of generating many such exercises, the tutor model will have many more sentences to choose from, and therefore will be able to choose exercises that are much better adapted to the vocabulary knowledge of a particular user.

6.2.4 Modelling user errors

The current system provides feedback, but the user model is informed only by the dictionary look-ups and the final outcome of the exercise, not by the details of the session. If we use the model only during one or a few sessions, this is not a major concern. For example, when a word is missing, the

feedback message provides a link to the dictionary. If the word is not looked up, and the exercise is eventually successfully completed, we can assume that the user has known this word before, and can regard it as learnt.

However, if we want to model user knowledge over larger periods of time, it is important to take such errors into account, even if the exercise is eventually successfully completed. The word that was used without any error is likely to be in user's long-term memory. The word that had to be elicited through feedback, even if it is known, probably has much lower activation level, and requires more repetition.

6.2.5 Improved sentence selection

The tutor model that selected exercises is relatively simple: the selection depends on utilities that are determined only by the user level and the number of unknown learning targets in the sentence.

An important insight from Vygotsky is that the zone of proximal development is not completely determined by the learner's current developmental level, but also depends on the individual. For example, assume that we have a group of learners on an identical level, with identical vocabulary knowledge, and an exercise with a large number of words that they do not know. The current system can either conclude that this exercise lies within the ZPD or outside the ZPD, but the conclusion will be same for all of them, while in fact, some of them may be determined enough to solve this exercise with the help of a dictionary, and some may not. Making the system learn such individual differences would probably be a good way to adapt to the user's knowledge and preferences even further.

There are also other ways of finding whether a sentence is suitable for a learner at a given level. Features that can be used to measure readability of sentences by second language learners at different levels were proposed and evaluated, among others, for Swedish (Pilán, Volodina & Johansson 2014) and French (François & Fairon 2012). In this system we are concerned with writability, so it is not clear whether all these features are relevant, and to what degree they can be used in other languages. They provide in any case a starting point for further investigation.

Appendix A

Exercises

This appendix presents the exercises that were created as a part of this thesis. They consist of answer patterns and error rules. The answer patterns allow the storing of a large number of answers in a compact manner – some exercises have over 1000 possible correct answers. The initial patterns and rules were based on information about Chinese grammar and common mistakes of English-speaking students provided by Robson (2012), Cheng (2009), Herzberg & Herzberg (2010) and the Chinese Grammar Wiki (Pasden et al. 2014). The sentence patterns and error rules were then verified and improved in collaboration with 3 native Chinese speakers. After testing and data collection the final adjustments were performed, based on actual user interactions with the system.

The exercises are presented in the following format: first, the English sentence is given. Then, a list of answer patterns follows, one per line. Patterns contain terminals (concrete words) and non-terminals; each of the non-terminals can be substituted by one of several synonymous expressions.

The non-terminals have English names that indicate their meaning, and are always surrounded by brackets. The non-terminals with names beginning with “o.” are optional – they may be substituted by a specific construction, or by an empty string. The non-terminals that have “m.” in their names represent classifiers, for example [m.buildings] is a classifier for buildings. The actual expressions that can appear in place of the non-terminals can be looked up in appendix B. Note that the non-terminals are sometimes smaller than expressions that are the learning targets. For example, there is no non-terminal for the whole “if..., then...” construction, but there are non-terminals for its components, e.g. [o.if2] for 的话 *dehuà*.

The last, optional part is the list of error patterns. The left column contains error patterns in the form of strings that are matched against user input (with ^ matching the beginning of the sentence, and / expressing alternative), and the right column contains the error message that is given if the match is successful.

1. We have breakfast at home

我们[at home]吃[breakfast]	
有	有 means ‘to have’ in the sense ‘to possess’. You need to say 吃 here.

2. My wife had lunch at a fast-food restaurant yesterday

我[wife] [yesterday] 在[o.m.fastfood] [fastfood] 吃了[lunch]	
[yesterday] 我[wife] 在[o.m.fastfood] [fastfood] 吃了[lunch]	
我[wife] [o.shi] [yesterday] 在[o.m.fastfood] [fastfood] 吃的[lunch]	
有	有 means 'to have' in the sense 'to possess'. You need to say 吃 here.
我的	When talking about relatives ('my sister', 'my wife', etc.), you shouldn't use 的. Using 的 in these cases may imply lack of closeness with those relatives.
昨日	昨日 is very formal and used exclusively in writing, you should use 昨天 instead
中餐	The primary meaning of 中餐 is 'Chinese food', to express 'lunch' it is better to say 午饭 or 午餐
内人	内人 does mean 'wife', but is rarely used. It is better to use a more common term.
夫人	This term does mean 'wife', but is quite formal and used only in some specific contexts. In a typical informal setting it is better to use terms such as 老婆 or 媳妇
爱人	This term does mean 'wife', but is quite formal and used by older generations. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇
妻子	This term does mean 'wife', but is quite formal. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇
太太	This term does mean 'wife', but is a bit formal. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇
一个	You can use 个 as a classifier for fast-food restaurants, but the sentence will sound better if you use a classifier for businesses

3. We've been living in Shanghai for 8 years

我们住[o.1e] 在上海[8] 年了	
我们在上海住[o.1e] [8] 年了	

4. He is as tall as I am

他[and] 我[same] 高	
我[and] 他[same] 高	
比	You can use 比 only if things are different

5. He is as busy as I am

他[and] 我[same] 忙	
我[and] 他[same] 忙	
比	You can use 比 only if things are different

6. I have a headache

我[headache]	
有	When describing a state, you shouldn't use 有.
了	With 了 the meaning becomes something more like 'I got a headache', emphasising the change of state. When you simply want to say that you have a headache now, without specifying whether you had it before, you shouldn't use 了

7. They all live in their native place, too

他们也[all] 住在[o.their] [native place]	
他们也[all] 在[o.their] [native place] 住	
都也/全也	The word for 'also, too' must be placed before the word for 'all'.
本土	本土 is one's native country, a 'native place' is 老家 or 家乡
本地	本地 means 'this place', a 'native place' is 老家 or 家乡
故乡	故乡 can only be used of one's native place that one left long time ago. A general term for a native place that you should use here is 老家 or 家乡.
本乡	本乡 is a rare word, used only in writing. A general term for a native place that you should use here is 老家 or 家乡.
籍贯	籍贯 is an official term for one's first registered place of living (户口). A general term for a native place that you should use here is 老家 or 家乡.

8. They all are Japanese, too

他们也[all] 是日本人	
都也/全也	The word for 'also, too' must be placed before the word for 'all'.

9. I have caught a cold

我[catch a cold]了	
有	When describing a state, you shouldn't use 有.

10. He is busy

他很忙	
是忙	他是忙 means 'he is in fact really busy', as a response to somebody else's comment that he might not be busy. In a neutral expression 'he is busy', you shouldn't use 是.

11. English is difficult

[English language]很难[o.learn]	
是难	英语是难 means 'English is in fact hard', as a response to somebody else's comment that it might not be hard. In a neutral expression 'English is hard', you shouldn't use 是.

12. Learning to do makeup is not difficult

[learn]化妆不难	
化装	化装 is the make-up that actors wear. When you write about make-up that women normally wear, you should use 化妆.
结构	结构 means 'make-up' in the sense of 'composition, make-up, the structure of things'. It is not the make-up that women normally wear
不很难	不很难 means 'not too difficult', to say 'not difficult' you should omit 很
不是很难	不是很难 means 'is not too difficult', to say 'not difficult' you should say 不很
不是难	不是难 is incorrect. You should omit 是.

13. These books are expensive

这些[books]很[expensive]	
是	With 是, the sentence means 'books are in fact expensive', as a response to somebody else's comment that they might be not. In a neutral expression 'books are expensive', you shouldn't use 是.
书籍	书籍 is a very formal term. You should rather use a more neutral and more common word, 书.
本	You shouldn't use a classifier, such as 本, when you are not giving a specific count.

14. Clothes are cheap

衣服很[cheap]	
衣服是	衣服是便宜 means 'clothes are in fact cheap', as a response to somebody else's comment that they might be not. In a neutral expression 'clothes are cheap', you shouldn't use 是.

15. I am extremely tired

我[extremely before adj]累	
我累[extremely after adj]	
我太累了	
累死我了	
很累极了	You cannot combine 很 with 极了
极为	极为 does mean 'extremely', but it quite rarely used. It's better to say 非常, 特别, 相当, 极其 before the adjective, or 极了 after the adjective

16. He runs extremely fast

他跑得[extremely before adj]快	
他跑得快[extremely after adj]	
他跑得太快了	
很快极了	You cannot combine 很 with 极了
极为	极为 does mean 'extremely', but it quite rarely used. It's better to say 非常, 特别, 相当, 极其 before the adjective, or 极了 after the adjective

17. That piano is extremely cheap

那[m.piano]钢琴[extremely before adj][cheap] 那[m.piano]钢琴[cheap][extremely after adj] 那[m.piano]钢琴太[cheap]了	
很便宜极了	You cannot combine 很 with 极了
极为	极为 does mean 'extremely', but it quite rarely used. It's better to say 非常, 特别, 相当, 极其 before the adjective, or 极了 after the adjective

18. There are three fast-food restaurants near the bank

[o.zai] 银行[o.de]附近有[3] [m.fastfood] [fastfood] 有[3] [m.fastfood] [fastfood] 在银行[o.de]附近 [o.zai] 银行附近[o.de] [fastfood] 有[3] [m.fastfood]	
三个/3个	You can use 个 as a classifier for fast-food restaurants, but the sentence will sound better if you use a classifier for businesses or a classifier for rooms
三座/3座	座 can be used for restaurants, but only if they are in separate buildings and you want to emphasise it. In this sentence it is much better to use a classifier for businesses or a classifier for rooms

19. There are two compact discs in the dictionary

[o.zai] [dictionary] [inside] [is inside] [2] 张[CD] 有[2] 张[CD] 在[dictionary] [inside]	
二	When counting items, you should use 两 instead of 二.
个	个 is not a very good choice for counting CDs. You should better use a classifier for flat objects
片	片 may be used for counting pages and leaves, for counting CDs it's not a best choice. You should better use a classifier for flat objects
内	内 as a word for 'inside' is quite formal, it is better to use 里 here

20. There are many people in the cafe

[o.zai] [cafe] [inside] 有[many] [o.de] 人 [o.zai] [cafe] [inside] [o.de] 人[many] 有[many] [o.de] 人在[cafe] [inside]	
有多/人多/人多	You need to add 很 before 多
一家咖啡/一间咖啡	You are talking about a specific cafe ('the cafe'), so you shouldn't use a count and a classifier before it
内	内 as a word for 'inside' is quite formal, it is better to use 里 here

21. There is a department store in this skyscraper

[o.zai] 这[m.buildings] [skyscraper] [inside] 有[o.m.department store] [department store] 有[o.m.department store] [department store] 在这[m.buildings] [skyscraper] [inside]	
摩天/高楼/高层建筑	It is a word for 'skyscraper', but it's better to use much more common terms, such as 大厦 or 大楼
所	所 can be used as a classifier only for relatively small buildings, not skyscrapers. 座 will work much better here
这个	个 is not a good measure word for buildings. 座 will work much better here
内	内 as a word for 'inside' is quite formal, it is better to use 里 here

22. That singer lives in this skyscraper

那[m.people] [singer] 住在这[m.buildings] [skyscraper] [o.inside] 那[m.people] [singer] 在这[m.buildings] [skyscraper] [o.inside] 住	
摩天/高楼/高层建筑	It is a word for 'skyscraper', but it's better to use much more common terms, such as 大厦 or 大楼
歌唱家	歌唱家 is not a term for singers in general, but only for those who sing exceptionally well
所	所 can be used as a classifier only for relatively small buildings, not skyscrapers. 座 will work much better here
这个	个 is not a good measure word for buildings. 座 will work much better here
内	内 as a word for 'inside' is quite formal, it is better to use 里 here

23. There are two cafes near the cinema

[o.zai] [cinema] [o.de] 附近有两 [m.cafe] [cafe] 有两 [m.cafe] [cafe] 在 [cinema] [o.de] 附近 [o.zai] [cinema] 附近 [o.de] [cafe] 有两 [m.cafe]	
在	The construction with 在 can be used only if we are talking about specific cafes ('the two cafes are ...'). This sentence contains 'two cafes', in indefinite form, so you need to say something like 'cinema - vicinity - there are - two cafes'.
二	When counting items, you should use 两 instead of 二.
两座	座 can be used for cafes, but only if they are in separate buildings and you want to emphasise it. In this sentence it is much better to use a classifier for businesses or a classifier for rooms
两个	You can use 个 as a classifier for cafes, but the sentence will sound better if you use a classifier for businesses or a classifier for rooms

24. There are many skyscrapers in this place

[o.at] 这 [o.m.place] [place] 有 [many] [o.de] [skyscraper] [o.at] 这 [o.m.place] [place] [skyscraper] [many] [o.de]	
有多	You need to add 很 before 多
摩天/高楼/高层建筑	It is a word for 'skyscraper', but it's better to use much more common terms, such as 大厦 or 大楼

25. He sold his piano yesterday

他 [yesterday] 卖了 [his own] 钢琴 [yesterday] 他卖了 [his own] 钢琴 他 [yesterday] 把 [his own] 钢琴卖了 [yesterday] 他把 [his own] 钢琴卖了	
昨日	昨日 is very formal and used exclusively in writing, you should use 昨天 instead
钢琴了	You may put 了 at the end, but in this kind of sentence it is more common to have it after the verb. Please place 了 after the verb ('sell')

26. He will sell his computer tomorrow

他 [tomorrow] [o.will] 卖 [his own] [computer] [tomorrow] 他 [o.will] 卖 [his own] [computer]	
明日	明日 is very formal and used exclusively in writing, you should use 明天 instead
翌日/旦日	翌日 is a name used in classical Chinese, you should use 明天 instead
将于	The construction with 将于 is very formal and used only in writing, so you should omit 于

27. There is a library across from the bank

[o.at] 银行 [o.de] 对面有一 [m.library] [library] 有一 [m.library] [library] 在银行 [o.de] 对面	
一个	You can use 个 as a classifier for libraries, but the sentence will sound better if you use a classifier for businesses
一座	座 can be used for libraries, but only if they are in separate buildings and you want to emphasise it. In general, it is much better to use a classifier for businesses
有图书馆	有图书馆 may mean 'there is a library' or 'there are libraries'. To be specific that you mean 'there is a library', you should add a number before 'library'.

28. I study at the library every evening

我 [every evening] 在 [library] 学习 [every evening] 我在 [library] 学习	
一个	This is *the* library, so you should not use 一个 ('a')
一家	This is *the* library, so you should not use 一家 ('a')
看书	看书 may mean 'to study', but its primary meaning is 'to read books'. Since library is a typical place for reading books, when you actually mean 'to study' in such context, you should use another word.
读书	读书 may mean 'to study', but only in a broad sense, such as 'study at a university' or 'attend a school'. You should use another word here.
念书	念书 may mean 'to study', but only in a broad sense, such as 'study at a university' or 'attend a school'. You should use another word here.

29. Every evening he reads books at home

他[every evening]在[at home][read books] [every evening]他在[at home][read books]	
念书	念书 means 'to read aloud a book', therefore it is not often used when somebody actually reads a book. It is more often used in its extended meaning: 'study at a university' or 'attend a school'

30. He works at the bank every day for 8 hours

他[every day]在银行[work][8][hours] [every day]他在银行[work][8][hours] 他[every day]在银行上[8][hours][o.de]班 [every day]他在银行上[8][hours][o.de]班	
一个	This is *the* bank, so you should not use 一个 ('a')
一家	This is *the* bank, so you should not use 一家 ('a')

31. I work at the cinema every evening for 3 hours

我[every evening]在[cinema][work][3][hours] [every evening]我在[cinema][work][3][hours] 我[every evening]在[cinema]上[3][hours][o.de]班 [every evening]我在[cinema]上[3][hours][o.de]班	
一个	This is *the* cinema, so you should not use 一个 ('a')
一家	This is *the* cinema, so you should not use 一家 ('a')

32. He came an hour earlier than I did

他比我早来[o.le][1][hours]	
来早/来了早	The word showing degree (早) must be placed before the verb (来)

33. He is 3 years older than me

他比我大[3]岁	
年岁	You should use 岁 alone, which is used for counting years of age.
年	年 means 'year', but here you need to use 岁 instead, which is used for counting years of age.

34. My computer is not as fast as yours

我的[computer]没有你的[o.computer][o.so]快	
没有快	没有 should be placed between the items that you are comparing
没有那么快	没有 should be placed between the items that you are comparing
不	To express 'is not as', you need to use 没有 instead of 不.

35. This compact disc is not as cheap as that one

这张[CD]没有那张[o.CD][o.so]便宜	
没有便宜/没有那么便宜	没有 should be placed between the items that you are comparing
不	To express 'is not as', you need to use 没有 instead of 不.
个	个 is not a very good choice for counting CDs. You should better use a classifier for flat objects
片	片 may be used for counting pages and leaves, for counting CDs it's not a best choice. You should better use a classifier for flat objects

36. I speak Chinese

我会[speak][Chinese language]	
我说	我说 may mean 'I am speaking' or 'I speak'. To be more precise, you should say something like 'I can speak Chinese'. This is the most common way to say it.
我讲	我讲 may mean 'I am speaking' or 'I speak'. To be more precise, you should say something like 'I can speak Chinese'. This is the most common way to say it.
能	能 means 'to have the possibility'. Here you need a word that means 'to know how to'.
可以	可以 means 'to be allowed to'. Here you need a word that means 'to know how to'.

37. He speaks Spanish

他会 [speak] [Spanish language]	
他说	他说 may mean 'he is speaking' or 'he speaks'. To be more precise, you should say something like 'he can speak Spanish'. This is the most common way to say it.
他讲	他讲 may mean 'he is speaking' or 'he speaks'. To be more precise, you should say something like 'he can speak Spanish'. This is the most common way to say it.
能	能 means 'to have the possibility'. Here you need a word that means 'to know how to'.
可以	可以 means 'to be allowed to'. Here you need a word that means 'to know how to'.

38. He speaks Chinese better than I do

他 [speak] [Chinese language] [speak] 得比我好 他比我 [speak] [Chinese language] [speak] 得好 他 [speak] [Chinese language] 比我 [speak] 得好	
会	In a construction for comparing verbs (the verb here is 'speak'), the verbs cannot be preceded by other verbs, such as 会.
很	When making comparisons, you shouldn't use 很.
非常	When making comparisons, you shouldn't use 非常.

39. He can read books faster than I do

他 [read] 书 [read] 得比我快 他比我 [read] 书 [read] 得快 他 [read] 书 比我 [read] 得快	
会	In a construction for comparing verbs (the verb here is 'read'), the verbs cannot be preceded by other verbs, such as 会.
很	When making comparisons, you shouldn't use 很.
非常	When making comparisons, you shouldn't use 非常.

40. None of us knows if the Chinese exam is difficult

我们都不知道 [Chinese language] [exam] 难不难	
要是/如果/若	To express 'if', as in 'if/whether it is true or not', you need to use V-不-V construction.

41. I don't know if he is tired

我不知道他累不累	
要是/如果/若	To express 'if', as in 'if/whether it is true or not', you need to use V-不-V construction.

42. I don't know if my younger sister is busy

我不知道我 [younger sister] 忙不忙	
要是/如果/若	To express 'if', as in 'if/whether it is true or not', you need to use V-不-V construction.
我的妹	When talking about relatives ('my sister', 'your mother', etc.), you shouldn't use 的. Using 的 in these cases may imply lack of closeness with those relatives.

43. My wife doesn't have a passport

我 [wife] [doesn't have] 护照	
不	When negating verbs such as 有, you need to use 没 instead of 不.
我的	When talking about relatives ('my sister', 'my wife', etc.), you shouldn't use 的. Using 的 in these cases may imply lack of closeness with those relatives.
内人	内人 does mean 'wife', but is quite rarely used. It is better to use a more common term.
夫人	This term does mean 'wife', but is quite formal and used only in some specific contexts. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇
爱人	This term does mean 'wife', but is quite formal and used by older generations. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇
妻子	This term does mean 'wife', but is quite formal. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇
太太	This term does mean 'wife', but is a bit formal. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇

44. If I had money, I would buy a house in my village

[o.if] 我有钱 [o.if2] , [if-then] 在我 [o.de] [village] 买房子 我 [o.if] 有钱 [o.if2] , [if-then] 在我 [o.de] [village] 买房子 [o.if] 有钱 [o.if2] , 我 [if-then] 在我 [o.de] [village] 买房子	
就我	就 needs to be placed after the subject, 我
买家	家 means 'home, family', but not 'house'. 'House' is 房子.
购买	购买 is a formal term. You should rather say 买
房屋	房屋 is a formal term. You should rather say 房子
农村	农村 means 'countryside', not an individual village. To say 'village', use 村子 or 村儿
乡下	乡下 means 'countryside', not an individual village. To say 'village', use 村子 or 村儿
乡村	乡村 is a formal term for a place recognised by authorities as a rural area. To say 'village', use 村子 or 村儿
村落	村落 is a very rare term. It is much better to say 村子 or 村儿
聚落	聚落 is a very rare term. It is much better to say 村子 or 村儿
村庄	村庄 is a term used mostly in poetry. It is much better to say 村子 or 村儿
在我里	里 is a Classical Chinese word for 'village', right now mostly used in names, where it may mean 'neighbourhood'. A Modern Chinese term for 'village' is 村子 or 村儿
在我的里	里 is a Classical Chinese word for 'village', right now mostly used in names, where it may mean 'neighbourhood'. A Modern Chinese term for 'village' is 村子 or 村儿
内	内 as a word for 'inside' is quite formal, it is better to use 里 here

45. If you're coming, call me or send me a text message

[o.if] 你来 [o.if2] , [o.then] [call me] [or] [text me] 你 [o.if] 来 [o.if2] , [o.then] [call me] [or] [text me]	
发我	You should write 给我发 instead of 发我
打我	You should write 给我打 instead of 打我
还是	还是 can only be used in questions, when a choice is presented. 'Or' in statements is 或者.

46. If I catch a cold, I'll return home

[o.if] 我感冒 [o.le] [o.if2] , 就回家 我 [o.if] 感冒 [o.le] [o.if2] , 就回家 [o.if] 感冒 [o.le] [o.if2] , 我就回家	
就我	就 needs to be placed after the subject, 我
有	When describing a state, you shouldn't use 有.

47. Not only does he speak Chinese, but he also speaks Spanish

他 [not only] 会 [speak] [Chinese language] , [but also] [o.too] 会 [speak] [Spanish language] 他又会 [speak] [Chinese language] , 又会 [speak] [Spanish language]	
不但他/不仅他/不只他	Both parts of the sentence have the same subject (他), therefore it must be placed before 不但.
能	能 means 'to have the possibility'. Here you need a word that means 'to know how to'.
可以	可以 means 'to be allowed to'. Here you need a word that means 'to know how to'.

48. Not only can he sing, but he also can play piano

他 [not only] 会唱歌 , [but also] [o.too] 会弹钢琴 他又会唱歌 , 又会弹钢琴	
不但他/不仅他/不只他	Both parts of the sentence have the same subject (他), therefore it must be placed before 不但.
能	能 means 'to have the possibility'. Here you need a word that means 'to know how to'.
可以	可以 means 'to be allowed to'. Here you need a word that means 'to know how to'.

49. I'm not only busy, but also tired

我[not only]很忙，[but also][o.I][o.too]很累 我又忙又累	
是忙	是忙 means 'in fact really busy', as a response to somebody else's comment that he might not be busy. In a neutral expression 'busy', you shouldn't use 是.
是累	是累 means 'in fact really tired', as a response to somebody else's comment that he might not be tired. In a neutral expression 'tired', you shouldn't use 是.

50. Although the library has many books, it has few dictionaries

[o.although][library]有[many]书，[dictionary]却[not many] [o.although][library]有[many]书，[but][dictionary][not many] [o.although][library]书[many]，[dictionary]却[not many] [o.although][library]书[many]，[but][dictionary][not many]	
，却	却 needs to be placed after the subject (dictionaries)
有多	You need to add 很 before 多
书籍	书籍 is a very formal term. You should rather use a more neutral and more common word, 书.
内	内 as a word for 'inside' is quite formal, it is better to use 里 here

51. Although my parents live in Beijing, I live in Hong Kong

[o.although]我[parents][o.all]住在北京，我却住在香港 [o.although]我[parents][o.all]住在北京，[but]我住在香港	
，却	却 needs to be placed after the subject (I)
我的	When talking about relatives ('my sister', 'my parents', etc.), you shouldn't use 的. Using 的 in these cases may imply lack of closeness with those relatives.

52. My younger sister texted me while I was reading a book

[o.just at]我[younger sister][send me a message][time]，我[in the middle of doing sth][read books]	
我的妹	When talking about relatives ('my sister', 'your mother', etc.), you shouldn't use 的. Using 的 in these cases may imply lack of closeness with those relatives.
念书	念书 means 'to read aloud a book', therefore it is not often used when somebody actually reads a book. It is more often used in its extended meaning: 'study at a university' or 'attend a school'
一边	You cannot use 一边, because there are two different subjects ('my sister' and 'I')
看一本书	看一本书 is a bit superfluous, 看书 sounds better.

53. My parents called me when I was having lunch

[o.just at]我[parents][call me][time]，我[in the middle of doing sth]吃[lunch]	
我的父母	When talking about relatives ('my sister', 'my parents', etc.), you shouldn't use 的. Using 的 in these cases may imply lack of closeness with those relatives.
一边	You cannot use 一边, because there are two different subjects ('my parents' and 'I')
有	有 means 'to have' in the sense 'to possess'. You need to say 吃 here.
中餐	The primary meaning of 中餐 is 'Chinese food', to express 'lunch' it is better to say 午饭 or 午餐

54. He was watching a film when I came

[o.just at]我[arrive][time]，他[in the middle of doing sth]看电影	
一边	You cannot use 一边, because there are two different subjects ('he' and 'I')

55. I work while having lunch

我[same time]吃[lunch]，[same time][work] 我[same time][work]，[same time]吃[lunch]	
时	时/的时候 pattern can only be used if there are two different subjects. There is only one subject here ('I'), so you need to use another construction.
有	有 means 'to have' in the sense 'to possess'. You need to say 吃 here.
中餐	The primary meaning of 中餐 is 'Chinese food', to express 'lunch' it is better to say 午饭 or 午餐

56. While having breakfast, I listen to music

我[same time]吃[breakfast], [same time]听音乐 我[same time]听音乐, [same time]吃[breakfast]	
时	时/的时候 pattern can only be used if there are two different subjects. There is only one subject here ('I'), so you need to use another construction.
有	有 means 'to have' in the sense 'to possess'. You need to say 吃 here.

57. This woman listens to music while applying makeup

这[o.m.people][woman][same time]化妆, [same time]听音乐 这[o.m.people][woman][same time]听音乐, [same time]化妆	
时	时/的时候 pattern can only be used if there are two different subjects. There is only one subject here ('I'), so you need to use another construction.
女人	For some people, 女人 has a negative connotation, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
妇女	妇女 is a quite formal term, and for some people it has a negative connotation, you should rather say 女的 (you may also say 女孩儿/女孩子 is the woman is relatively young)
女子	女子 is a term used mainly in poetry, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
小姐	You should be careful with using 小姐, as in some situations it may mean 'prostitute', it is safer to say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
女性	女性 is a quite formal term, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
化装	化装 is the make-up that actors wear. When you write about make-up that women normally wear, you should use 化妆.
结构	结构 means 'make-up' in the sense of 'composition, make-up, the structure of things'. It is not the make-up that women normally wear

58. First I will finish my breakfast, then I will go watch a film

[o.I]先吃完[breakfast], [then][o.I][o.go/then]看电影	
以后	In 'first...then' construction, you need to use 然后, giving equal stress to both actions. When using 以后, the sentence will sound more like 'After finishing breakfast, I'll go watch a film', with emphasis on the second action.
后来	后来 is used when two last for a long time and are not tightly one after another. In 'first...then' construction, you should rather use 然后.
首先	首先 is a very formal word, it is better to simply say 先

59. First I will go to Shanghai, then I will go to Beijing

[o.I]先[go to]上海, [then][o.I][o.then][go to]北京 [o.I]先到上海去, [then][o.I][o.then]到北京去	
以后	In 'first...then' construction, you need to use 然后, giving equal stress to both actions. When using 以后, the sentence will sound more like 'after going to Shanghai, I'll go to Beijing', with emphasis on the second action.
后来	后来 is used when two last for a long time and are not tightly one after another. In 'first...then' construction, you should rather use 然后.
首先	首先 is a very formal word, it is better to simply say 先

60. First I'm going to the cinema, then I'm going to listen to music

[o.I]先[go to][cinema], [then][o.I][o.go/then]听音乐 [o.I]先到[cinema]去, [then][o.I][o.go/then]听音乐	
以后	In 'first...then' construction, you need to use 然后, giving equal stress to both actions. When using 以后, the sentence will sound more like 'After going to the cinema, I'm going to listen to music', with emphasis on the second action.
后来	后来 is used when two last for a long time and are not tightly one after another. In 'first...then' construction, you should rather use 然后.
首先	首先 is a very formal word, it is better to simply say 先

61. He went to Hong Kong again

他又[go to]了香港 他又[go to]香港了 他又到香港去了	
再	The action happened in the past, so you need to use 又 instead of 再.

62. As soon as you finish watching this film, I will call you

你[o.as soon as]看完这[m.film]电影，我就[call you]	
就我	就 needs to be placed after the subject ('I')

63. Can he sing?

他会唱歌吗？ 他会不会唱歌？	
会不会唱歌吗	To make a question, you should either use V不V pattern, or end the sentence with 吗, but not both.
能	能 means 'to have the possibility'. Here you need a word that means 'to know how to'.
可以	可以 means 'to be allowed to'. Here you need a word that means 'to know how to'.

64. What music is on this CD?

这张[CD]上有什么音乐？	
个	个 is not a very good choice for counting CDs. You should better use a classifier for flat objects
片	片 may be used for counting pages and leaves, for counting CDs it's not a best choice. You should better use a classifier for flat objects

65. Where do you like going to buy clothes?

你喜欢[go/at][where]买衣服？ 你喜欢到[where]去买衣服？	
---	--

66. What music do you like?

你喜欢[o.listen]什么音乐？	
--------------------	--

67. Where does that Japanese live?

那[m.people]日本人住在[where]？ 那[m.people]日本人在[where]住？	
--	--

68. This girl is 20 years old

这[o.m.people][woman][20][years of age]	
年	年 means 'year', but here you need to use 岁, which is used for counting years of age.
年岁	You should use 岁 alone, which is used for counting years of age.
女人	For some people, 女人 has a negative connotation, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
妇女	妇女 is a quite formal term, and for some people it has a negative connotation, you should rather say 女的 (you may also say 女孩儿/女孩子 is the woman is relatively young)
女子	女子 is a term used mainly in poetry, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
小姐	You should be careful with using 小姐, as in some situations it may mean 'prostitute', it is safer to say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
女性	女性 is a quite formal term, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)

69. Where is my passport?

[my]护照在[where]？	
-----------------	--

70. Many women like to buy clothes

[many] [o.de] [woman] 喜欢买衣服	
女人	For some people, 女人 has a negative connotation, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
妇女	妇女 is a quite formal term, and for some people it has a negative connotation, you should rather say 女的 (you may also say 女孩儿/女孩子 is the woman is relatively young)
女子	女子 is a term used mainly in poetry, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
小姐	You should be careful with using 小姐, as in some situations it may mean 'prostitute', it is safer to say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
女性	女性 is a quite formal term, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)

71. My younger sister likes to use makeup

我 [younger sister] 喜欢化妆	
我的	When talking about relatives ('my sister', 'my parents', etc.), you shouldn't use 的. Using 的 in these cases may imply lack of closeness with those relatives.
化装	化装 is the make-up that actors wear. When you write about make-up that women normally wear, you should use 化妆.
结构	结构 means 'make-up' in the sense of 'composition, make-up, the structure of things'. It is not the make-up that women normally wear

72. Do you know if he is at home this evening?

你知道不知道他今天晚上在不在 [at home] ? 你知道他今天晚上在不在 [at home] 吗 ?	
要是/如果/若	To express 'if', as in 'if/whether it is true or not', you need to use V-不-V construction.
这	It is incorrect to use 这 to express 'this evening', you should write something like 'today in the evening'

73. I don't know how much that Chinese dictionary sells for

我不知道那 [m.dictionary] [Chinese language] [dictionary] 卖多少钱	
---	--

74. Next summer I will learn English or Spanish

[next year] 夏天我 [o.will] [learn] [English language] 或者 [Spanish language] 我 [next year] 夏天 [o.will] [learn] [English language] 或者 [Spanish language]	
还是	还是 can only be used in questions, when a choice is presented. 'Or' in statements is 或者.
下年	The correct term for 'next year' is 明年
下	下 is not a typical Chinese way to express 'next' in 'next summer', you should rather write something like 'the coming year in the summer'
暑天	暑天 means 'summer day', to express 'summer' you need to write 夏天
暑假	暑假 means 'summer vacation', to express 'summer' you need to write 夏天
岁	岁 is a year of age. A calendar year is 年.
夏季	夏季 is a formal term for 'summer'. You should rather say 夏天
过年	过年 means 'to spend the Spring Festival'. To say 'next year', you should rather say 明年
将于	The construction with 将于 is very formal and used only in writing, so you should omit 于

75. He will be 30 years old next summer

[next year] [summer] 他 [30] [years of age] 他 [next year] [summer] [30] [years of age]	
30年/三十年	年 means 'year', but here you need to use 岁, which is used for counting years of age.
下年	The correct term for 'next year' is 明年
下	下 is not a typical Chinese way to express 'next' in 'next summer', you should rather write something like 'the coming year in the summer'
暑天	暑天 means 'summer day', to express 'summer' you need to write 夏天
暑假	暑假 means 'summer vacation', to express 'summer' you need to write 夏天
夏季	夏季 is a formal term for 'summer'. You should rather say 夏天
过年	过年 means 'to spend the Spring Festival'. To say 'next year', you should rather say 明年

76. They are going to marry next summer

[next year] [summer] 他们 [o.will] 结婚 他们 [next year] [summer] [o.will] 结婚 他们将于 [next year] [summer] 结婚	
下年	The correct term for 'next year' is 明年
下	下 is not a typical Chinese way to express 'next' in 'next summer', you should rather write something like 'the coming year in the summer'
暑天	暑天 means 'summer day', to express 'summer' you need to write 夏天
暑假	暑假 means 'summer vacation', to express 'summer' you need to write 夏天
岁	岁 is a year of age. A calendar year is 年.
夏季	夏季 is a formal term for 'summer'. You should rather say 夏天
过年	过年 means 'to spend the Spring Festival'. To say 'next year', you should rather say 明年
下年	The correct term for 'next year' is 明年

77. Tomorrow I will read books or watch films

[tomorrow] 我 [o.will] [read books] [or] [o.watch] 电影 [tomorrow] 我 [o.will] 看电影 [or] [o.watch] 书 我 [tomorrow] [o.will] [read books] [or] [o.watch] 电影 我 [tomorrow] [o.will] 看电影 [or] [o.watch] 书 我将于 [tomorrow] 看电影 [or] [o.watch] 书	
还是	还是 means 'or' only in questions, when a choice is presented. 'Or' in statements is 或者.
明日	明日 is very formal and used exclusively in writing, you should use 明天 instead
翌日	翌日 is a name used in classical Chinese, you should use 明天 instead
旦日	旦日 is a name used in classical Chinese, you should use 明天 instead
念书	念书 means 'to read aloud a book', therefore it is not often used when somebody actually reads a book. It is more often used in its extended meaning: 'study at a university' or 'attend a school'

78. Is that an American passport or a Chinese one?

那是 [o.m.passport] 美国护照还是 [o.m.passport] 中国护照? 那 [m.passport] 护照是美国 [o.de] 还是中国的?	
或	或者 means 'or' only in statements, not in questions. 'Or' in questions is 还是.
中国?	中国 means 'China'. To say 'Chinese', you need to add 的 at the end.

79. Is that a fast-food restaurant or a cafe?

那是 [cafe] 还是 [fastfood] ? 那是 [fastfood] 还是 [cafe] ?	
或	或 means 'or' only in statements, not in questions. 'Or' in questions is 还是.

80. I'm not often hungry before eating lunch

[o.I] 吃[lunch] [before] [o.comma] [o.I] 不常饿	
的前/饭前/餐前	You cannot use 前 alone to say 'before' here. Use a two-character word, such as 以前, that has the same meaning.
不常常	To express 'not often' simply say 不常
不经常	不经常 is OK, but it's better to simply say 不常

81. When did you arrive in Hong Kong?

你[o.shi] 什么时候[arrive] 香港的?	
你[o.shi] 什么时候到香港来的?	
了	This is a question about details (time) of a past action that is already known (arrive in Hong Kong). In this situation you shouldn't use 了, but rather use the (是)...的-construction. Please put 的 at the end of the sentence.
香港?/来?	This is a question about details (time) of a past action that is already known (arrive in Hong Kong). You should therefore use the (是)...的-construction. Please put 的 at the end of the sentence.

82. In which year did they get married?

他们[o.shi] [which] 年结婚的?	
了	This is a question about details (time) of a past action that is already known (getting married). In this situation you shouldn't use 了, but rather use the (是)...的-construction. Please put 的 at the end of the sentence.
结婚?	This is a question about details (time) of a past action that is already known (getting married). The sentence must therefore end with 的 (and the subject 'they' may be followed by 是).

83. How did you return home?

你[o.shi] [how] 回家的?	
你怎么回家?	你怎么回家? means 'how will (you) return home'. You need a question that refers to the past
你怎么回了家?/你怎么回家了?	Your sentence means 'how come you returned home?', which may be asked when you are surprised that someone went home. To express 'How did you return home?', you need to use (是)...的. Please put 的 at the end of the sentence.

84. I'm not returning home until tomorrow

我[tomorrow] 才回家	
[tomorrow] 我才回家	
才明	才 should be placed after the specification of time ('tomorrow')
不/没	In Chinese, to express 'not until', you shouldn't use negation, but rather the word 才, which means 'at that time (but not before)'.
明日	明日 is very formal and used exclusively in writing, you should use 明天 instead
翌日	翌日 is a name used in classical Chinese, you should use 明天 instead
旦日	旦日 is a name used in classical Chinese, you should use 明天 instead

85. These five students are Japanese

这[5] [m.student] 学生是日本人	
这些	When giving a specific number of items, you should use 这, not 这些.

86. Those three books are mine

那[3] [m.book] [books] 是我的	
那些	When giving a specific number of items, you should use 那, not 那些.
个	The classifier for books is 本, not 个.

87. These 2 people are singing

[these people] [2 people] [in the middle of doing sth] 唱歌	
这些	When giving a specific number of items, you should use 这, not 这些.
二	When specifying a number of something, you should use 两 instead of 二.
人唱歌/位唱歌/俩唱歌	To specify that they are singing right now (and not just sing in general), please add a word that means 'in the middle of doing something'

88. These 2 people are singers

[these people] [2 people] 是 [singer]	
这些	When giving a specific number of items, you should use 这, not 这些.
二	When specifying a number of something, you should use 两 instead of 二.
歌唱家	歌唱家 is not a term for singers in general, but only for those who sing exceptionally well

89. I often send him text messages

我 [often] [send him messages]	
-------------------------------	--

90. My parents often watch films

我 [parents] [often] 看电影	
我的	When talking about relatives ('my sister', 'my parents', etc.), you shouldn't use 的. Using 的 in these cases may imply lack of closeness with those relatives.

91. My younger sister often sings

我 [younger sister] [often] 唱歌	
我的	When talking about relatives ('my sister', 'my parents', etc.), you shouldn't use 的. Using 的 in these cases may imply lack of closeness with those relatives.

92. They often speak English

他们 [often] 说 [English language]	
---------------------------------	--

93. I don't often catch cold

我不常感冒	
了	The sentence is about having cold in general, so you shouldn't use 了, which is used only if there is some change of state
不常常/不太	To express 'not often' simply say 不常
不经常	不经常 is OK, but it's better to simply say 不常
有	When describing a state, you shouldn't use 有.

94. He doesn't often buy computers

他不常买电脑	
了	The sentence is about buying computers in general, so you shouldn't use 了, which is used only if there is some change of state
不常常/不太	To express 'not often' simply say 不常
不经常	不经常 is OK, but it's better to simply say 不常

95. As soon as he divorced his wife, he married another woman

他[o.as soon as][with]他[o.de][wife]离了婚，就[with][another][woman]结了婚	
他[o.as soon as][with]他[o.de][wife]离了婚，就娶了[another][woman]	
结婚了	The verb here is 结, so you should write 结了婚
离婚了	The verb here is 离, so you should write 离了婚
离婚，	You need a completed action marker 了 to show that 离婚 happened in the past
离了婚他	To express 'divorce a person' you need to write something like 'with a person get divorced'.
结了婚他	To express 'marry a person' you need to write something like 'with a person get married'.
内人	内人 does mean 'wife', but is quite rarely used. It is better to use a more common term.
夫人	This term does mean 'wife', but is quite formal and used only in some specific contexts. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇
爱人	This term does mean 'wife', but is quite formal and used by older generations. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇
妻子/太太	This term does mean 'wife', but is quite formal. In a typical informal setting it is better to use other terms, such as 老婆 or 媳妇
妇女	妇女 is a quite formal term for 'woman', in a typical informal context you should use another one
女士	女士 is a quite formal term for 'woman', in a typical informal context you should use another one
女子	女子 is a term used mainly in poetry, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
小姐	You should be careful with using 小姐, as in some situations it may mean 'prostitute', it is safer to say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)
女性	女性 is a quite formal term, you should rather say 女的 (you may also say 女孩儿/女孩子/姑娘 is the woman is relatively young)

96. That place is remote/off the beaten track/out-of-the-way

那[o.m.place][place][remote]	
那[o.m.place][place]人迹罕至	

97. The native place of this singer is in a remote, out-of-the-way village

这[m.people][singer]的[native place]在[o.m.village][remote]的[village]	
歌唱家	歌唱家 is not a term for singers in general, but only for those who sing exceptionally well
本土	本土 is one's native country, a 'native place' is 老家 or 家乡
本地	本地 means 'this place', a 'native place' is 老家 or 家乡
本乡	本乡 is a rare word, used only in writing. A general term for a native place that you should use here is 老家 or 家乡.
籍贯	籍贯 is an official term for one's first registered place of living (户口). A general term for a native place that you should use here is 老家 or 家乡.
农村	农村 means 'countryside', not an individual village. To say 'village', use 村子 or 村儿
乡下	乡下 means 'countryside', not an individual village. To say 'village', use 村子 or 村儿
乡村	乡村 is a formal term for a place recognised by authorities as a rural area. To say 'village', use 村子 or 村儿
村落	村落 is a very rare term. It is much better to say 村子 or 村儿
聚落	聚落 is a very rare term. It is much better to say 村子 or 村儿
村庄	村庄 is a term used mostly in poetry. It is much better to say 村子 or 村儿
在里	里 is a Classical Chinese word for 'village', right now mostly used in names, where it may mean 'neighbourhood'. A Modern Chinese term for 'village' is 村子 or 村儿
的里	里 is a Classical Chinese word for 'village', right now mostly used in names, where it may mean 'neighbourhood'. A Modern Chinese term for 'village' is 村子 or 村儿

98. Where is your native place?

你[o.de] [native place] 在[where] ?	
本土	本土 is one's native country, a 'native place' is 老家 or 家乡
本地	本地 means 'this place', a 'native place' is 老家 or 家乡
本乡	本乡 is a rare word, used only in writing. A general term for a native place that you should use here is 老家 or 家乡.
籍贯	籍贯 is an official term for one's first registered place of living (户口). A general term for a native place that you should use here is 老家 or 家乡.

99. Is there any village nearby?

[here]附近有[village]吗?	
[here]附近有没有[village]?	
[here]附近有[village]没有?	
农村	农村 means 'countryside', not an individual village. To say 'village', use 村子 or 村儿
乡下	乡下 means 'countryside', not an individual village. To say 'village', use 村子 or 村儿
乡村	乡村 is a formal term for a place recognised by authorities as a rural area. To say 'village', use 村子 or 村儿
村落	村落 is a very rare term. It is much better to say 村子 or 村儿
聚落	聚落 is a very rare term. It is much better to say 村子 or 村儿
村庄	村庄 is a term used mostly in poetry. It is much better to say 村子 or 村儿
有里	里 is a Classical Chinese word for 'village', right now mostly used in names, where it may mean 'neighbourhood'. A Modern Chinese term for 'village' is 村子 or 村儿

100. I didn't see him yesterday

我[yesterday] [did not] [see] 他	
[yesterday] 我 [did not] [see] 他	
[yesterday] 我 [did not] [with] 他见面	
我[yesterday] [did not] [with] 他见面	
了	When negating actions in the past, you shouldn't use 了
看他	看 means to 'look at'. You need to write 看见 to make it mean 'see'.
见面他	To express 'meet a person', you need to write something like 'with a person meet'.
不	When negating verbs such as 有, you need to use 没 instead of 不.
昨日	昨日 is very formal and used exclusively in writing, you should use 昨天 instead

Appendix B

Synonyms

[1] ::= 一 1	[every evening] ::= 每晚 每天晚上 天天晚上
[2] ::= 两 2	[exam] ::= 考试 测试
[3] ::= 三 3	[expensive] ::= 贵 昂贵
[2 people] ::= 两个人 两位 俩	[extremely after adj] ::= 极了 坏了 死了 死我了
[5] ::= 五 5	[extremely before adj] ::= 十分 真 非常 特别 相当 极其
[8] ::= 八 8	[fastfood] ::= 快餐店 速食店 快餐厅
[10] ::= 十 10	[go/at] ::= 去 到 在
[20] ::= 二十 20	[go to] ::= 去 到
[30] ::= 三十 30	[headache] ::= 头疼 头痛
[all] ::= 都 全 全都	[here] ::= 这儿 这里 这
[and] ::= 和 跟 同	[his own] ::= 他 自己 他自己 他的 自己的
[another] ::= 别的 另一个 其他 另外一个	[hours] ::= 个小时 个钟头 小时
[arrive] ::= 到 来 来到	[house] ::= 房子 房 屋 屋子
[at home] ::= 在家里 在家	[how] ::= 怎么 怎 咋
[at home] ::= 家 家里	[if-then] ::= 就 就会 就要 会 要
[before] ::= 以前 之前	[inside] ::= 里 里面 里头
[books] ::= 书	[in the middle of doing sth] ::= 正在 在
[breakfast] ::= 早饭 早餐	[is inside] ::= 有 夹着
[but also] ::= 而且 并且 同时	[learn] ::= 学 学习 学会
[but] ::= 但是 可是	[library] ::= 图书馆 圉
[cafe] ::= 咖啡店 咖啡屋 咖啡厅 咖啡馆 咖啡馆儿	[lunch] ::= 中饭 午饭 午餐
[call me] ::= 给我打电话 打电话给我 打电话 打你电话 打给你电话	[many] ::= 很多 不少
[call you] ::= 给你打电话 打电话给你 打你电话 打给你电话	[m.book] ::= 本
[catch a cold] ::= 感冒 伤风 着凉	[m.buildings] ::= 座 栋
[CD] ::= 光盘 光碟	[m.cafe] ::= 家 间
[cheap] ::= 便宜 不贵 廉价	[m.dictionary] ::= 本 部
[Chinese language] ::= 汉语 中文 中国话 普通话 国语	[m.fastfood] ::= 家 间
[cinema] ::= 电影院 影院	[m.film] ::= 个 部 片
[computer] ::= 电脑 计算机	[m.library] ::= 家
[department store] ::= 百货商场 百货商店 百货公司 百货店 杂货行	[m.passport] ::= 个 本
[dictionary] ::= 字典 辞典 词典	[m.people] ::= 个 位
[did not] ::= 没 没有	[m.piano] ::= 架 台
[doesn't have] ::= 没有 没	[m.student] ::= 个 位 名
[English language] ::= 英语 英文	[my] ::= 我 我的
[every day] ::= 每天 天天	[native place] ::= 家乡 老家 故乡
	[next year] ::= 明年 来年 未来的一年
	[not many] ::= 不多 很少

[not only] ::= 不但 不仅 不只	爸妈妈 爸妈
[o.all] ::= 都	[place] ::= 地方 地区
[o.although] ::= 虽然	[read books] ::= 看书 读书
[o.as soon as] ::= 一	[read] ::= 看 读 念
[o.at] ::= 在	[remote] ::= 偏僻 很偏僻 很遥远 很背 人迹罕至
[o.CD] ::= 光盘 光碟 碟片	[same time] ::= 一边 一边儿
[o.comma] ::= ,	[same] ::= 一样 同样
[o.computer] ::= 电脑 计算机	[see] ::= 看见 见到 见 看到
[o.de] ::= 的	[send him messages] ::= 给他发短信 发短信给他 发给他短信 发他短信
[often] ::= 常常 经常 往往	[send me a message] ::= 给我发短信 发短信给我 发给我短信 发我短信
[o.go/then] ::= 就 再 去 就去 再去	[singer] ::= 歌手 演唱者 歌者
[o.if] ::= 如果 要是	[skyscraper] ::= 大楼 大厦
[o.if2] ::= 的话	[Spanish language] ::= 西班牙语 西班牙文 西语 西文
[o.inside] ::= 里 里面	[speak] ::= 说 讲
[o.I] ::= 我	[summer] ::= 夏天
[o.just at] ::= 当	[text me] ::= 发短信 给我发短信 发短信给我 发给我短信 发我短信
[o.learn] ::= 学	[then] ::= 然后
[o.le] ::= 了	[these people] ::= 这 他们
[o.listen] ::= 听	[time] ::= 的时候 时
[o.m.department store] ::= 家 间 个 一家 一间 一个	[tomorrow] ::= 明天 明儿
[o.m.fastfood] ::= 一家 一间	[village] ::= 村子 村儿
[o.m.passport] ::= 一个 一本	[where] ::= 哪儿 哪里 哪
[o.m.people] ::= 个 位	[which] ::= 哪 哪一
[o.m.place] ::= 个 处 块	[wife] ::= 老婆 媳妇
[o.m.village] ::= 一个 一座	[with] ::= 跟 和
[or] ::= 或者 或	[woman] ::= 姑娘 女的 女士 女孩儿 女孩 女孩子
[o.shi] ::= 是	[work] ::= 工作 干活
[o.so] ::= 那么 这么	[years of age] ::= 岁 岁了 了
[o.their] ::= 他们 他们的	[yesterday] ::= 昨天
[o.then] ::= 就 再	[younger sister] ::= 妹妹 妹子 小妹
[o.too] ::= 还 也	
[o.watch] ::= 看	
[o.will] ::= 要 会 将	
[o.zai] ::= 在	
[parents] ::= 父母 父母亲 老人家 爸	

Bibliography

- Aleven, Vincent. 2010. Rule-based cognitive modeling for intelligent tutoring systems. In Roger Nkambou, Jacqueline Bourdeau & Riichiro Mizoguchi (eds.), *Advances in intelligent tutoring systems*, 33–62. Berlin, Heidelberg: Springer-Verlag.
- Beckwith, Richard, Christiane Fellbaum, Derek Gross & George A. Miller. 1991. WordNet: A lexical database organized on psycholinguistic principles. *Lexical acquisition: Exploiting on-line resources to build a lexicon*. 211–232.
- Bourdeau, Jacqueline & Monique Grandbastien. 2010. Modeling tutoring knowledge. In Roger Nkambou, Jacqueline Bourdeau & Riichiro Mizoguchi (eds.), *Advances in intelligent tutoring systems*, 123–143. Springer.
- Bravo, Héctor Corrada & Raghu Ramakrishnan. 2007. Optimizing MPF queries: Decision support and probabilistic inference. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, 701–712. ACM.
- Cheng, Meizhen [程美珍]. 2009. 汉语病句辨析九百例: *Error analysis of 900 sample sentences*. Zhu Li [李珠] (ed.). Sinolingua.
- Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2(3). 113–124.
- Conati, Cristina. 2010. Bayesian student modeling. In *Advances in intelligent tutoring systems*, 281–299. Springer.
- Corbett, Albert T. & John R. Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4(4). 253–278.
- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Tech. rep. Cambridge.
- Da, Jun. 2004. A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In *Proceedings of the fourth International Conference on New Technologies in Teaching and Learning Chinese*, 501–11. Beijing: Tsinghua University Press.
- Díez, Francisco J. & Marek J. Druzdzel. 2006. *Canonical probabilistic models for knowledge engineering*. Tech. rep. Technical Report CISIAD-06-01, UNED, Madrid, Spain.
- Dubois, Daniel, Roger Nkambou, Jean-François Quintal & François Savard. 2010. Decision-making in cognitive tutoring systems. In Roger Nkambou,

- Jacqueline Bourdeau & Riichiro Mizoguchi (eds.), *Advances in intelligent tutoring systems*, 145–179. Springer.
- Ebbinghaus, Hermann. 1885/1913. *Memory: A contribution to experimental psychology*. Original work published 1885. New York: Teachers College, Columbia University.
- Evans, Norman W., K. James Hartshorn, Robb M. McCollum & Mark Wolfersberger. 2010. Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research* 14(4). 445–463.
- François, Thomas & Cédric Fairon. 2012. An AI readability formula for French as a foreign language. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 466–477. Association for Computational Linguistics.
- Goldberg, Adele. 2006. *Constructions at work*. Oxford: Oxford University Press.
- Gong, Yue, Joseph E. Beck & Neil T. Heffernan. 2010. Comparing knowledge tracing and performance factor analysis by using multiple model fitting procedures. In *Intelligent tutoring systems*, 35–44. Springer.
- Hartshorn, K. James, Norman W. Evans, Paul F. Merrill, Richard R. Sudweeks, Diane Strong-Krause & Neil J. Anderson. 2010. Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly* 44(1). 84–109.
- Hayden, Jeffrey J. 2005. Breaking the camel’s back: Cognitive load and reading Chinese. In *International Interdisciplinary Conference on Hanzi renzhi – How Western learners discover the world of written Chinese*. Gemersheim, Germany.
- Hayes, Edmund B. 1987. The relationship between Chinese character complexity and character recognition. *Journal of the Chinese Language Teachers Association* 22(2). 45–57.
- Heift, Trude & Devlan Nicholson. 2001. Web delivery of adaptive and interactive language tutoring. *International Journal of Artificial Intelligence in Education* 12(4). 310–325.
- Herzberg, Qin Xue & Larry Herzberg. 2010. *Basic patterns of Chinese grammar: A student’s guide to correct structures and common errors*. Stone Bridge Press.
- ISO, International Organization for Standardization. 1992. *Information Technology – Database Language SQL*. Tech. rep. (Second Informal Review Draft) ISO/IEC 9075:1992.
- Jokinen, Kristiina. 2009. *Constructive dialogue modelling: Speech interaction and rational agents*. Vol. 11. Wiley-Interscience.
- Jokinen, Kristiina & Michael McTear. 2009. Spoken dialogue systems. *Synthesis Lectures on Human Language Technologies* 2(1). 1–151.
- Jurafsky, Daniel & James H. Martin. 2008. *Speech and language processing (Prentice Hall Series in Artificial Intelligence)*. Prentice Hall.
- Koda, Keiko. 2005. *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.
- Koehn, Philipp. 2009. *Statistical machine translation*. Cambridge University Press.

- Koller, Daphne & Nir Friedman. 2009. *Probabilistic graphical models: Principles and techniques*. MIT press.
- Kosek, Michał. in press. Output-based vocabulary teaching software with interactive feedback. In *Proceedings of the 6th International Symposium on SLA in China*.
- Kosek, Michał & Pierre Lison. in press. An intelligent tutoring system for learning Chinese with cognitive model of the learner. In *2014 EUROCALL Proceedings*.
- Krashen, S. 1985. *The input hypothesis*. London: Longman.
- Lantolf, James P. & Steven L. Thorne. 2007. Sociocultural theory and second language learning. *Theories in second language acquisition: An introduction*. 201–224.
- Laufer, Batia. 1997. The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. *Second language vocabulary acquisition*. 20–34.
- Laurillard, Diana & Giorgio Marullo. 1993. Computer-based approaches to second language learning. In *Language, classrooms and computers*, 145–165. Routledge.
- Leonard, David. 2002. *Learning theories: A to Z*. Greenwood.
- Levenshtein, Vladimir I. 1966. Binary codes capable of correcting deletions, insertions and reversals. In *Soviet physics doklady*, vol. 10, 707.
- Levy, Mike & Glenn Stockwell. 2006. *Call dimensions: Options and issues in computer-assisted language learning*. Mahwah, NJ: Routledge.
- Lewis, Michael. 1993. *The lexical approach*. Vol. 1. Language Teaching Publications Hove.
- Ling, Vivian. 2007. Studies on L2 acquisition of the Chinese script published in america. *The Cognition, Learning and Teaching of Chinese Characters*. 51–83.
- Lyster, Roy & Kazuya Saito. 2010. Oral feedback in classroom SLA. *Studies in Second Language Acquisition* 32(02). 265–302.
- Ma, Qing & Peter Kelly. 2006. Computer assisted vocabulary learning: Design and evaluation. *Computer Assisted Language Learning* 19(1). 15–45.
- Mace, Cecil Alec. 1968. *The psychology of study*. Penguin.
- McCalla, Gordon. 2010. May the Forcing Functions be with You. In Roger Nkambou, Jacqueline Bourdeau & Riichiro Mizoguchi (eds.), *Advances in intelligent tutoring systems*, VII–X. Berlin, Heidelberg: Springer-Verlag.
- Nkambou, Roger. 2010. Modeling the domain: An introduction to the expert module. In Roger Nkambou, Jacqueline Bourdeau & Riichiro Mizoguchi (eds.), *Advances in intelligent tutoring systems*, 15–32. Berlin, Heidelberg: Springer-Verlag.
- Nkambou, Roger, Jacqueline Bourdeau & Riichiro Mizoguchi. 2010. *Advances in intelligent tutoring systems*. Vol. 308. Berlin & Heidelberg: Springer-Verlag.
- Ohlsson, Stellan. 1996. Learning from performance errors. *Psychological review* 103(2). 241.

- Olney, Andrew M., Arthur C. Graesser & Natalie K. Person. 2010. Tutorial dialog in natural language. In Roger Nkambou, Jacqueline Bourdeau & Riichiro Mizoguchi (eds.), *Advances in intelligent tutoring systems*, 181–206. Berlin, Heidelberg: Springer-Verlag.
- Ortega, Lourdes. 2013. *Understanding second language acquisition*. London & New York: Routledge.
- Papineni, Kishore, Salim Roukos, Todd Ward & Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on Association for Computational Linguistics*, 311–318. Association for Computational Linguistics.
- Pasden, John et al. 2014. *Chinese Grammar Wiki*. Retrieved 2014-07-30. <http://resources.allsetlearning.com/chinese/grammar/>.
- Pass, Susan. 2004. *Parallel paths to constructivism: Jean Piaget and Lev Vygotsky*. IAP.
- Pavlik, Philip I. 2006. *The microeconomics of learning: Optimizing paired-associate memory*. ProQuest Information & Learning PhD thesis.
- Pavlik, Philip I. & John R. Anderson. 2010. Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science* 29(4). 559–586.
- Pearl, Judea. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.
- Peyton Jones, Simon et al. 2003. The Haskell 98 language and libraries: The revised report. *Journal of Functional Programming* 13(1). 1–255. <http://www.haskell.org/definition/>.
- Pilán, Ildikó, Elena Volodina & Richard Johansson. 2014. Rule-based and machine learning approaches for second language sentence-level readability. *ACL 2014*. 174.
- Reye, Jim. 1998. Two-phase updating of student models based on dynamic belief networks. In *Intelligent tutoring systems*, 274–283. Springer.
- Robson, Shin Yong. 2012. *Speed up your Chinese: Strategies to avoid common errors*. Routledge.
- Russell, Stuart Jonathan & Peter Norvig. 2010. *Artificial intelligence: A modern approach*. Prentice hall Upper Saddle River, NJ.
- Schmidt, R. 1983. Interaction, acculturation, and the acquisition of communicative competence. In N. Wolfson & E. Judd (eds.), *Sociolinguistics and language acquisition*, 137–74. Rowley, MA: Newbury House.
- Schmidt, R. 1995. Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (ed.), *Attention and awareness in foreign language learning*, 1–63. Honolulu, HI: National Foreign Language Resource Center.
- Selwyn, Neil. 2011. *Education and technology: Key issues and debates*. Bloomsbury Publishing.
- shànghuǒ, 上火. 2009. *CC-CEDICT*. Chinese–English dictionary. Retrieved 2014-07-30. http://cc-cedict.org/editor/editor.php?log_id=9138&return=ListChanges&handler=ViewLogEntry.
- Shen, Helen. 2007. Vocabulary acquisition and reading comprehension: Evidence from advanced non-native Chinese learners. *The cognition, learning and teaching of Chinese characters*.

- Swain, M. 1985. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (eds.), *Input in second language acquisition*, 235–253. Rowley, MA: Newbury House.
- Tharp, Roland G., Ronald Gallimore, et al. 1988. *Rousing minds to life*. Cambridge: Cambridge University Press.
- Thornbury, Scott. 2006. *An A–Z of ELT: A dictionary of terms and concepts used in English language teaching*. Macmillan Education.
- Thornbury, Scott. 2013. *Big questions in ELT*. the round.
- Truscott, John. 2007. The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing* 16(4). 255–272.
- de Vries, Bart P., Catia Cucchiarini, Stephen Bodnar, Helmer Strik & Roeland van Hout. 2014. Spoken grammar practice and feedback in an ASR-based CALL system. *Computer Assisted Language Learning*. 1–27.
- Vygotsky, Lev S. 1978. *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, Lev S. [Выготский Лев С.]. 1935. *Умственное развитие детей в процессе обучения* [Mental development of children in the process of education]. Russian. Moscow & Leningrad: Gosudarstvennoe uchebno-pedagogicheskoe izdatel'stvo. <http://psychlib.ru/mgppu/VUR/VUR-1935.html>.
- Wong, Kam-Fai, Wenjie Li, Ruifeng Xu & Zheng-sheng Zhang. 2009. Introduction to Chinese natural language processing. *Synthesis Lectures on Human Language Technologies* 2(1). 1–148.
- Wong, S.K. Michael, Yang Xiang & Xiaopin Nie. 1995. Representation of Bayesian networks as relational databases. In *Advances in intelligent computing – IPMU'94*, 117–130. Springer.
- Wood, David, Jerome S. Bruner & Gail Ross. 1976. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry* 17(2). 89–100.
- Woolf, Beverly Park. 2010. Student modeling. In Roger Nkambou, Jacqueline Bourdeau & Riichiro Mizoguchi (eds.), *Advances in intelligent tutoring systems*, 267–279. Berlin, Heidelberg: Springer-Verlag.
- Woźniak, Piotr A. & Edward J. Gorzelańczyk. 1994. Optimization of repetition spacing in the practice of learning. *Acta Neurobiologiae Experimentalis* 54. 59–62.
- Xiong, Hui [熊卉]. 2006. 从主题与主语看英汉句子结构特点 [Features of English sentence structure from the perspective of topic and subject]. Chinese. 吉首大学学报：社会科学版 [Journal of Jishou University: Social Science] 5. 031.
- Zong, Chengqing [宗成庆]. 2007. 统计自然语言处理 [Statistical natural language processing]. Chinese. Beijing: Tsinghua University Press.