



HAL
open science

Building a Learning Space for Language Acquisition

Eduardo Calò

► **To cite this version:**

Eduardo Calò. Building a Learning Space for Language Acquisition. Computation and Language [cs.CL]. 2021. hal-03389328

HAL Id: hal-03389328

<https://inria.hal.science/hal-03389328>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE LORRAINE
IDMC

MASTER'S THESIS

Building a Learning Space for Language Acquisition

Author:
Eduardo CALÒ

Supervisor:
Prof. Miguel COUCEIRO

Co-supervisors:
Dr. Alexandre BAZIN
Dr. Benedict O'DONNELL

Reviewer:
Prof. Maxime AMBLARD

*A thesis submitted in fulfillment of the requirements
for the degree of Master of Science in Natural Language Processing*

in

Orpailleur, Loria
Naimrod, Inria



March 2021 - August 2021

Declaration of Authorship

I, Eduardo CALÒ, declare that this thesis titled, “Building a Learning Space for Language Acquisition” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:



Date: August 24, 2021

UNIVERSITÉ DE LORRAINE

Abstract

Master of Science in Natural Language Processing

Building a Learning Space for Language Acquisition

by Eduardo CALÒ

Learning spaces offer a rigorous mathematical foundation for algorithms that can be employed in intelligent tutoring systems to organize the body of knowledge necessary to become proficient in a domain. These mathematical structures are particular types of partial ordered sets, and thus they can be built using formal concept analysis (FCA), which provides powerful formal tools, such as concept lattices, to order data of this nature.

Learning spaces are helpful to bring to light implications between notions and uncover students' different learning paths. These structures would eventually help teachers create personalized learning plans and intelligently assess students. Learning spaces have been used to ease teaching and learning in a variety of subjects. However, applications in language acquisition remain challenging.

This manuscript presents an adaptation of this framework to language acquisition, specifically Mandarin, by leveraging the shared properties between learning space theory and FCA. The functioning of the tool realized in this work is based on a local and dynamic exploration of the space, without the need to compute the whole structure beforehand. This way, a scalable system architecture, which is not limited by the size of the input and the power of the machine on which it runs, is achieved.

We report experiments conducted using several datasets and feedback provided by teachers who used the first prototype of the system. The results we obtain are encouraging and suggest that the approach can be adapted to other languages.

Acknowledgements

During the past two years, many things have happened. I'd need an entire chapter to list everything. However, I don't have much space, so I'll try to be brief. But, let's go in order...

Before anything else, I would like to express my most profound appreciation to my supervisor Prof. Miguel Couceiro, and my co-supervisors, Dr. Alexandre Bazin and Dr. Benedict O'Donnell. Without your support, commitment, and feedback, this work would never have seen the light of day. Although it was a short ride, thank you all for letting me feel a pivot of the project.

Miguel, you are an awesome person and an invaluable guide, both inside and outside academia. Your knowledge is unparalleled, yet your humbleness is hard to find in this world. Your warm-hearted, joyful, and chill attitude (and your calls on Saturday evenings!) made me always feel at home. To me, you're more like a friend than a supervisor. Looking forward to celebrating this achievement with you!

Alexandre, your expertise has been infinitely precious to the advancement of my work. The scientific discussions we had about the algorithms to implement had a massive impact on steering this project in the right direction.

Ben, I'm deeply indebted to you for showing me how to be a great leader. With you on my side, I could grow professionally and mature personally. I got to know how entrepreneurship works and how to hold an interview. Ah, and let's not forget the constant tips you gave me in previous drafts of the thesis!

I also wish to thank my "Parrain" Maxime Amblard, who will take some of his time to read my thesis during his summer holidays. Hope you will enjoy reading it, anyway!

The (unfortunately remote) work environment has always been friendly and pleasant. I thank Inria for funding my internship and Marc, Paul, and Issei for being my loyal colleagues. I wish I could have had this journey together with you in person. Special thanks also go to the beta-testers and teachers who tried our prototypes and gave useful feedback. Finally, I would like to thank all the WenTest-takers around the world: without you, there would be no learning space. 谢谢大家，朋友们!

I would like to acknowledge our colleagues at the University of Helsinki, with whom we developed a fruitful collaboration. In particular, special thanks go to Prof. Roman Yangarber for our constructive discussions regarding the results obtained.

This whole ride's been a pretty tough one. The master's workload has always been enormous (even if the results have been gratifying!), and COVID made everything more difficult. Fortunately, I've constantly been flanked by lovely persons. My *vie nancéienne* wouldn't have been the same without my "*gros · se · s*". They've been my inseparable companions, especially during COVID times. The travels, picnics, and dinners, or the nights spent playing board games, greatly alleviated this challenging period. It's incredible how time is really relative. I feel like I've known them my whole life, but it's actually been just a couple of years, or even less in some cases.

Last but not least, I'm extremely grateful to the people who've decided (forcibly or otherwise!) to stick around me for years. First, my family that has always supported me in any choice of my life. I wouldn't be anywhere close to where I am now without their care (and my mum feeding me when I'm home!). Then, my long-time friends

that keep bearing me as the years pass by. You're almost family to me. Finally, my mentors who've accompanied me and helped in my choices throughout the course of my life. Without your inspiration, I would certainly be now doing something very different.

Well, now I'm done, for real! Another chapter of my life is over. I'm curious to know where I'll be heading next...

Contents

Abstract	v
Acknowledgements	vii
Contents	ix
List of Figures	xi
List of Tables	xiii
List of Abbreviations	xv
1 Introduction	1
1.1 General Framework	1
1.2 Context and Objective	3
2 Background	5
2.1 Knowledge Space Theory	5
2.2 Formal Concept Analysis	10
2.3 Linking KST and FCA	11
2.4 Language Assessment Criteria	12
2.5 Related Work	14
3 Implementation	17
3.1 Methodology	17
3.2 Data Acquisition	18
3.2.1 Selection of the Skills	18
3.2.2 WenTest	20
3.3 The Learning Space	21
3.3.1 Space Construction	22
3.3.2 Assessment Process	22
3.3.3 Learning Path	25
3.4 Integration into Naimrod's Framework	27
4 Experiments and Results	29
4.1 Lattice Construction	29
4.2 Mandarin Data	31
4.2.1 Analysis of the Structure	31
4.3 Russian Data	33
4.3.1 Testing the Assessment Process	35
4.4 Teachers' Opinions	36
4.5 Discussion	37

5 Conclusion	39
5.1 Difficulties	39
5.2 Future Work	40
5.3 Final Words	41
Bibliography	43
A List of Mandarin Skills	49
B Best Mandarin Path	59
C List of Russian Skills	67
D Best Russian Path	71

List of Figures

1	Surmise graph of the items in Table 1.	7
2	Learning space derived from the items in Table 1. The bold lines highlight the fringes of state $\{c, d\}$. The red lines indicate a possible learning path. These notions will be discussed in the following paragraphs.	8
3	Two examples of knowledge structures.	9
4	Example of formal concept analysis taken from Chaikalis, 2016.	12
5	CEFR levels' descriptors.	13
6	Distribution of the 195 skills with respect to their HSK tags.	19
7	WenTest's interface.	20
8	Visualization of a step of the assessment process: I is the introductory state, c in red is the skill with the median score in S , and the state highlighted in red is the updated knowledge state of the student after a correct response to an instance of c	24
9	Visualization of a step of the algorithm for the selection of the best learning path.	26
10	Learning Space's interface.	27
11	Knowledge structure built using the data gathered from WenTest, with the best learning path highlighted in red.	32

List of Tables

1	Sample knowledge domain, with its items and their instances.	6
2	Sample dataset in the form of formal context.	21
3	Experiments with several off-the-shelf tools. Columns values indicate the dimension of the input formal context. T = time error. M = memory error.	30
4	Full list of Mandarin skills with corresponding HSK and CEFR tags, ordered by increasing HSK level.	57
5	Best path identified in the knowledge structure in Figure 11 built from the data gathered from WenTest.	65
6	Full list of Russian skills with corresponding CEFR tags, ordered by increasing CEFR level.	70
7	Best path identified from Revita's data.	74

List of Abbreviations

ALEKS	A ssessment and L earning in K nowledge S paces
AOC-poset	A tttribute and O bject C oncept- p artially o rdered s et
BKT	B ayesian K nowledge T racing
BLIM	B asic L ocal I ndependence M odel
CALL	C omputer- A ided L anguage L earning
CEFR	C ommon E uropean F ramework of R eference for Languages
FCA	F ormal C oncept A nalysis
GRM	G raded R esponse M odel
HMM	H idden M arkov M odel
HSK	H anyu S huiping K aoshi
HSKK	H anyu S huiping K ouyu K aoshi
IRT	I tem R esponse T heory
ITS	I ntelligent T utoring S ystem
KST	K nowledge S pace T heory
MML	M arginal M aximum L ikelihood
NLP	N atural L anguage P rocessing

Chapter 1

Introduction

In the present thesis, we will describe the work that was conducted during the author's internship at Loria¹ and Naimrod², a recently born startup commercializing educational technology for language learning that is incubated at Inria³. The topic of the internship was building a learning space for Mandarin acquisition.

The following sections of this chapter will cover the general research framework in which this work can be collocated (Section 1.1) and the specific context that gave birth to the work of the internship (Section 1.2).

The rest of the thesis is structured as follows. Chapter 2 will present the theoretical notions needed to understand this thesis and discuss some related works. Then, Chapter 3 will detail the methodology and the implementation steps followed for the realization of the work. Next, Chapter 4 will present the results of the evaluations we carried out on the tool. Finally, the thesis will conclude with Chapter 5 discussing the difficulties encountered throughout the project and mentioning ideas to improve the work in the future.

1.1 General Framework

Asking computers to help language practitioners and scholars in the arduous task of learning and teaching a language dates back to the 1960s (Hart, 1981). Computer-aided language learning (CALL) (Levy, 1997) is a broad, multidisciplinary field of information technology precisely focused on the applications of computers to support professors and students with language teaching and learning.

CALL has been an active research area for years, but it is particularly in vogue nowadays, thanks to the recent improvements that we have seen in computational linguistics and natural language processing (NLP). Some of the most popular CALL tools commercially available at the moment are Babbel⁴, Duolingo⁵, and LingoDeer⁶.

Crucially, most of these tools provide all the users with a predetermined and fixed set of contents and exercises. Offering the same closed and limited set of exercises to everyone conflicts with the principles of adaptability of the learning process and established pedagogical practices, especially as regards languages.

¹<https://www.loria.fr/fr/>

²<http://www.naimrod.com>

³<https://www.inria.fr/en>

⁴<https://www.babbel.com/>

⁵<https://www.duolingo.com/>

⁶<https://www.lingodeer.com/>

In contrast to well-defined subjects, such as mathematics or physics, which are based on a clearly segmented formal theory and can be assessed through questions whose responses can be classified unambiguously as right or wrong, ill-defined subjects, such as languages, ethics, and music, do not have a clear-cut theory and questions presented to pupils are hardly definable as unambiguously right or wrong (Lynch et al., 2006). Thus, the progression path of a student of languages may vary considerably compared to the ones of their peers.

Although some CALL systems do track personal progress, future exercises are typically not adapted depending on users' prior errors and knowledge. In fact, the set of exercises presented is always the same for everyone.

Thus, in order to give the best possible experience to learners, CALL tools should conjointly work with intelligent tutoring systems (ITS). ITSs (Psootka et al., 1988) are computer systems whose role is to give adaptive, specific, and personalized lessons, exercises, and suggestions to students as a human tutor would.

Any ITS should include at least a *knowledge model*, a *student model*, and a *tutoring model* (Nkambou, Mizoguchi, and Bourdeau, 2010). The knowledge model is the component that encodes the notions that a student should gradually master to become proficient in the given subject. The student model tracks the users' progress and monitors how they interact with the system. Finally, the tutoring model is responsible for estimating the difficulty of the notions in the body of knowledge and deciding how to present them. This way, lessons and exercises are adapted to users' needs, and they are presented with appropriate content to avoid boredom or frustration.

Apart from simple heuristics, several theories can be implemented in an ITS to organize knowledge and provide users with a valuable experience. Among those, we can find knowledge space theory (KST) (Falmagne et al., 2013), item response theory (IRT) (Lord, 2012), and Bayesian knowledge tracing (BKT) (Corbett and Anderson, 1994). KST will receive an appropriate and lengthy introduction in Section 2.1 since it is the mathematical foundation of this thesis, while IRT and BKT are briefly outlined here.

IRT (Embretson and Reise, 2013) is a psychometric paradigm used to conceive, analyze, and rate instruments that measure the abilities or characteristics of individuals, such as tests and questionnaires. The theory is based on the relationship between individuals' responses to a test and the overall properties of the items measured in the test. Most importantly, IRT does not assume that each item is equally difficult. In fact, the difficulty of each item aims to be embedded as information in scoring items.

Essentially, IRT models the difficulty of items in a test and test-takers' ability, given several response patterns of individuals to the same test. Section 3.3 will present a description of a specific IRT model we employ in our system.

BKT is a widespread algorithm embedded in numerous ITSs. Specifically, it tries to model students' knowledge and problem sequencing in a hidden Markov model (HMM) (Baum et al., 1972), updated by tracking student interactions with the system over time. BKT assumes a dichotomous approach to skills' mastery and observations: a skill is either mastered by the student or not, and a student can only get a problem either right or wrong.

By monitoring students' performances over time, BKT can assess whether the student has successfully mastered the skills. More precisely, there are four parameters to account for in BKT:

1. the probability of the student knowing the skill before the first time they see it;
2. the probability of the student luckily guessing the correct application of an unknown skill;
3. the probability of the student accidentally making a mistake when applying a known skill;
4. the probability of the student demonstrating mastering of a skill after a single attempt.

Implementation of BKT in our system was out of scope of this internship. However, its integration in the future is not excluded.

1.2 Context and Objective

Naimrod's mission is to tailor language lessons to the level and interests of each learner. Their solution should help teachers cut the time needed to craft engaging lessons for their pupils.

The first language the startup is tackling is Mandarin. The choice of this language is motivated by the booming demand for Mandarin courses over the past decade (Zhang and Li, 2019) and the challenge it still presents. It typically takes Western learners more than 2,000 hours to reach a conversational level of Mandarin⁷. Because of this, students tend to lose interest in their courses, and dropout rates are high.

The internship's objective was to help Naimrod address these issues by creating a CALL service augmented by an ITS that incorporates KST, which would smartly order concepts to achieve fluency in Mandarin.

To reach this goal, the choice fell on constructing a learning space. Learning spaces have the advantage of bringing to light implications between notions and uncovering students' different learning paths and strategies. The mathematical structure would eventually help teachers create personalized learning plans, assess students quickly and frequently to identify gaps in their knowledge before they grow, and improve the overall teaching quality.

⁷<https://www.state.gov/foreign-language-training/>

Chapter 2

Background

The scope of this chapter is twofold: giving an overview of the theoretical background needed to understand this study and presenting related works. First, we will recall the mathematical background that was useful for the development of this tool. More precisely, in Section 2.1, we will discuss some key notions in knowledge space theory (KST), while Section 2.2 will be devoted to formal concept analysis (FCA). Then, in Section 2.3, we will establish some links between these two frameworks and, in Section 2.4, recall the linguistic assessment criteria that we found suitable to explore in our application. Finally, Section 2.5 will be dedicated to discussing some existing works pertaining to ours.

2.1 Knowledge Space Theory

Knowledge space theory (KST) (Falmagne and Doignon, 2010) is a framework in mathematical psychology, which originated from Doignon and Falmagne, 1985 to address the shortcomings of traditional assessment tests for students (such as the ACT⁸, SAT⁹, or typical state standards tests). The main issue with these tests is that their possible outcomes are too coarse. Indeed, they result in placing a student in one of a few dozen categories. These resulting broad categories are not sufficient to delineate the exact knowledge of a student effectively. On the other hand, KST presents a fine-grained approach to learners' assessment. In fact, the basic principle of KST is that of uncovering the exact knowledge (namely, a detailed set of concepts) mastered by a student.

In KST, an *item* is a discrete type of problem or skill that a student should master. A specific example of an item is called *instance*. Each instance of the same item should be carefully chosen to be equally difficult and cover the same content. The collection of all the items is called the *domain* of knowledge. A *knowledge state* is a subset of items in the domain that a student may know at a specific time. It is crucial to note that not any subset can be a knowledge state. The reasons for this will become clear in the following paragraphs. Finally, the collection of all the knowledge states forms a *knowledge space*.

To give an intuition of the notions mentioned above, we will provide a clarifying example, before formally defining them¹⁰. Table 1 presents a sample knowledge domain stemming from high school mathematics. The first column is the collection of

⁸<https://www.act.org/content/act/en/products-and-services/the-act.html>

⁹<https://collegereadiness.collegeboard.org/sat>

¹⁰The example is taken and adapted from Falmagne et al., 1990.

Item	Instance
a. Multiplication of large numbers	$378 \cdot 605 = ?$
b. Multiplication of decimal numbers	$58.7 \cdot 0.94 = ?$
c. Multiplication of fractions	$\frac{1}{2} \cdot \frac{5}{6} = ?$
d. Calculation of percentages	What is 30% of 34?
e. Problems with fractions	Gwendolyn is $\frac{3}{4}$ as old as Rebecca. Rebecca is $\frac{2}{5}$ as old as Edwin. Edwin is 20 years old. How old is Gwendolyn?

TABLE 1: Sample knowledge domain, with its items and their instances.

the items belonging to this knowledge domain, while the second column presents concrete instances of the items in the first column that would be presented to students when they are assessed.

To test pupils, one could assess mastery of all the items in the domain. However, such a method would be impractical with extensive domains. Efficiency can be improved by taking advantage of the connections between these problems. For example, any student mastering item e probably masters also items a , b , and c . Thus, assessing these three items would be a waste of time. Therefore, we can safely assume that items a , b , and c are prerequisites to item e . A reasonable knowledge state that can be derived from the domain is hence $\{a, b, c, e\}$. All the knowledge states created by following these inferences deriving from the responses given by pupils form a knowledge space.

Formally, a knowledge space is a pair (Q, \mathcal{K}) consisting of a nonempty set Q and a collection \mathcal{K} of subsets of Q , with $\emptyset \in \mathcal{K}$ (corresponding to a student who knows nothing of the subject) and $Q \in \mathcal{K}$ (corresponding to a student being fully competent in the subject). The set Q is called the domain of the knowledge space (Q, \mathcal{K}) . The elements of Q are the items, and the elements of \mathcal{K} are the knowledge states. A crucial property satisfied by knowledge spaces is that of being union-closed¹¹.

Notably, a knowledge space does not correspond to the powerset of Q since not any subset of Q can be a knowledge state. Indeed, from observing that a student masters a given item, one can sometimes infer that this student also masters other items. This suggests the introduction of a *surmise relation*, which is defined as follows: item q can be surmised from item t if and only if from a correct response to an instance of t we can infer a correct response to an instance of q . Thus, we can now precisely define a knowledge state as a subset K of Q that contains some item q and all the items that can be surmised from q .

At this point, we can say that knowledge spaces are a type of *surmise systems* (Doignon and Falmagne, 2012). This property reflects the practical fact that if a person masters an item q , then they master all items in at least one state in which q is present. This is consonant with the view that there may be more than one way (thus different sets of prerequisites) to achieve mastery of any particular item q . From this property, a surmise graph of items can be built. The surmise graph is a Bayesian

¹¹A collection of sets is *union-closed* if the union of any two sets from the collection belongs to the collection. Likewise, a collection of sets is *intersection-closed* if the intersection of any two sets from the collection belongs to the collection.

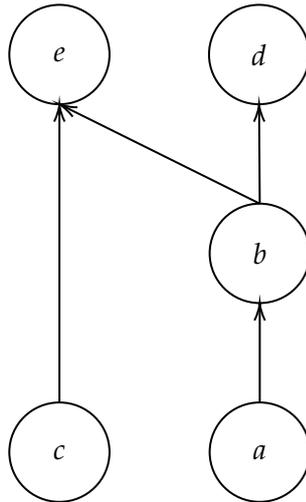


FIGURE 1: Surmise graph of the items in Table 1.

network (Pearl, 2009), with each node representing one item. The direction of the arcs between these nodes stands for their surmise relationship.

A plausible surmise graph for the five items in Table 1 is shown in Figure 1. From this representation, a complete set of the possible knowledge states can be inferred and then organized in a knowledge space. A graphical representation of the resulting knowledge space is given in Figure 2.

Several methods have been proposed in the literature to build a knowledge structure from scratch or derive it from a surmise graph. These methods can be grouped into three main categories (Chiusole, Stefanutti, and Spoto, 2017):

- *query to experts* (Müller, 1989; Koppen and Doignon, 1990; Dowling, 1993; Koppen, 1993; Kambouri et al., 1994; Schrepp and Held, 1995) consists of algorithms to build knowledge structures incrementally by posing questions to domain experts;
- *skill map construction* (Lukas and Albert, 1993; Doignon, 1994; Albert and Lukas, 1999; Heller et al., 2013) refers to the creation of a mapping between a set of items and the skills of an individual as a way to build knowledge structures;
- *data-driven approaches* (Villano, 1991; Schrepp, 1999; Schrepp, 2003; Sargin and Ünlü, 2009; Spoto, Stefanutti, and Vidotto, 2016) aim at directly deriving knowledge structures from data or questionnaires without any prior assumptions about the relations among items.

When a knowledge space satisfies two additional properties, a particular subcase arises. The particular structure that is obtained is denominated *learning space* and has the benefit of offering a more complete representation of the organization of the knowledge¹². The two properties that should be satisfied are called *learning smoothness* and *learning consistency*. The former refers to the fact that there should be no gaps in the learning material: an individual should be able to master items one by one. The latter implies that an advanced student should feel more at ease learning

¹²To note that the set Q of items in a learning space is finite.

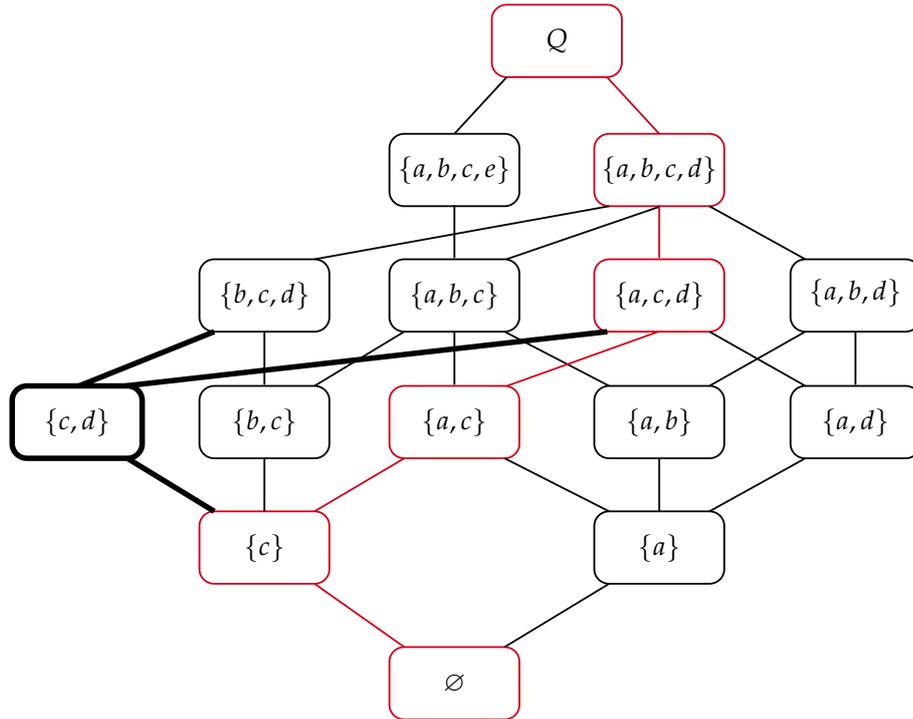


FIGURE 2: Learning space derived from the items in Table 1. The bold lines highlight the fringes of state $\{c, d\}$. The red lines indicate a possible learning path. These notions will be discussed in the following paragraphs.

a new item than another less competent student. These two properties are formally stated as follows.

LEARNING SMOOTHNESS: for any two knowledge states K, L with $K \subset L$, there exist a finite chain of knowledge states

$$K = K_0 \subset K_1 \subset \dots \subset K_p = L \quad (1)$$

such that $|K_i \setminus K_{i-1}| = 1$ for $1 \leq i \leq p$ (thus we have $|L \setminus K| = p$).

LEARNING CONSISTENCY: for any two knowledge states K, L with $K \subset L$, if q is an item such that $K \cup \{q\}$ is a knowledge state, then $L \cup \{q\}$ is also a knowledge state.

To clear things up, Figure 3 presents two knowledge structures on the same domain $Q = \{a, b, c, d\}$. The structure in Figure 3a is a knowledge space since it only satisfies learning consistency but not learning smoothness. Indeed, if we take states $\{c, d\}$ and Q , we have that (i) $\{c, d\} \subset Q$ and (ii) no other states exist between them. However, $|Q \setminus \{c, d\}| = 2$ since both a and b need to be added simultaneously to $\{c, d\}$ to obtain Q . Thus, learning smoothness is not achieved. On the other hand, the structure in Figure 3b is a learning space since it satisfies both learning smoothness and learning consistency, achieving a one-by-one transition between all connected states.

The structures presented in this section are somewhat simple. However, a knowledge state may potentially contain a vast number of items, thus making its representation quite cumbersome. Fortunately, another essential property of learning spaces allows the representation to be slender, employing just a meaningful summary in the

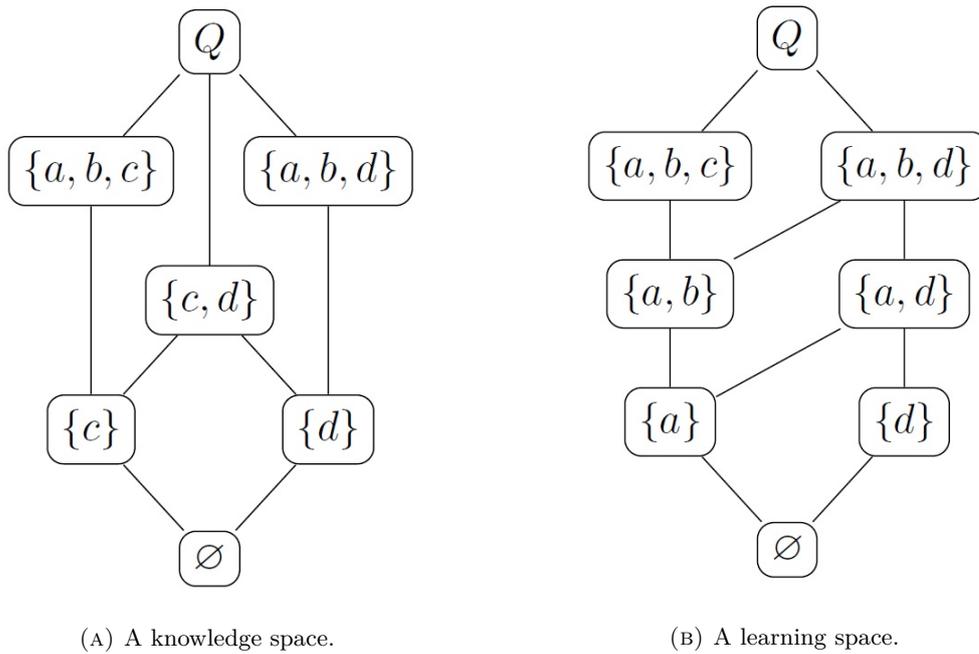


FIGURE 3: Two examples of knowledge structures.

form of *outer fringe* and *inner fringe*. The outer fringe of a knowledge state contains all the items that the student is ready to learn, so it is particularly significant under a pedagogical perspective. Conversely, the inner fringe contains all the items representing the actual peak of the student's competence. Formally, the outer fringe of a state K is the set of all the items q such that adding q to K yields another knowledge state, while the inner fringe of a state K is the set of all the items q such that removing q from K yields another knowledge state¹³. In the structure in Figure 2, the inner and outer fringes of state $\{c, d\}$ are highlighted. Specifically, the inner fringe corresponds to $\{d\}$, while the outer fringe to $\{a, b\}$. In this particular case, the summary may seem useless. However, in realistic learning spaces with millions of states, the summary may be considerably smaller than the state.

In a learning space, there are many ways to successfully acquire all the items in the domain. This property relates to the fundamental concept of *learning path*. A learning path is any sequence of states in which a student, starting from the empty state, could learn all the items and reach full proficiency in the domain. Since any realistic learning space typically contains millions of knowledge states, trillions of different learning paths can undoubtedly be found in such a space. In the learning space in Figure 2, a possible learning path is highlighted in red. The corresponding learning sequence is the following: c, a, d, b, e .

It is important to note that, as for the description given so far, the concept of knowledge structures is deterministic. However, since the theory's inception, it has been deemed fundamental to introduce probabilities to provide more realistic predictions of subjects' responses.

¹³It is noteworthy that the empty state always has an empty inner fringe, and that the whole domain Q has an empty outer fringe.

Probabilities are indeed fundamental within the last core element of KST: the *assessment process*. The original idea consists of an adaptive algorithm that gradually updates, after each response, the probability distribution over all the knowledge states in \mathcal{K} . After n questions being answered, the most probable state is considered the one that represents better the actual knowledge of the subject.

Several things should be considered when carrying out the assessment process. Each response is categorically considered to be either correct or incorrect. Moreover, the possibility of *lucky guesses* and *careless errors* by the subjects must be taken into account or eradicated (i.e., by removing multiple-choice questions). In fact, in some situations, a subject may guess the correct response to a question not yet mastered or may be careless in responding to a mastered item, making an error. Probabilistic models have been implemented to deal with these issues, with the most widely applied being the basic local independence model (BLIM) (Falmagne and Doignon, 1988; Doignon and Falmagne, 1999). Lastly, given that, in real applications, the number of states composing a knowledge space is in the order of millions, this assessment algorithm needs to be parallelized by splitting the space into several sub-spaces and applying the algorithm alternatively on those (Doignon and Falmagne, 2016).

2.2 Formal Concept Analysis

Formal concept analysis (FCA) (Ganter and Wille, 2012) is a field of applied mathematics based on concept discovery, classification, and exploration. It was developed in Wille, 1981 and builds on Birkhoff’s lattice theory (Birkhoff, 1940), which itself has its roots in set theory (Cantor, 1874). The original motivation behind FCA is the search for a method to conceptualize and structure real-world knowledge into a formal mathematical structure. Being a very versatile formalism, FCA has found multiple applications in many branches of science (Priss, 2006; Poelmans et al., 2013; Singh, Aswani Kumar, and Gani, 2016).

FCA is based on a representation of data into binary tables, denominated formal contexts. A *formal context* is a relational structure (G, M, I) , with G (from the German “Gegenstände”) a set of *objects*, M (from the German “Merkmale”) a set of *attributes*, and $I \subseteq G \times M$ a binary relation, called *incidence*, between G and M . An object $g \in G$ possesses an attribute $m \in M$, if gIm holds. Refer to Figure 4a for an exemplar formal context.

Between the objects and the attributes of a formal context, a *Galois connection* (Stewart, 2015) is established. For any set $A \subseteq G$ and $B \subseteq M$, the following two *derivation operators* define the Galois connection:

$$A' = \{m \in M \mid (g, m) \in I \forall g \in A\} \quad (2)$$

$$B' = \{g \in G \mid (g, m) \in I \forall m \in B\} \quad (3)$$

which are, respectively, the set of all attributes shared by all objects from A , and the set of all objects sharing all attributes from B .

Then, a *formal concept* is defined to be a pair (A, B) , where $A \subseteq G$, $B \subseteq M$, and $A' = B$ and $B' = A$. A is called the *extent* and B the *intent* of the concept.

Derivation operators can be composed. Applying either one of the two and then the other constitutes two *closure operators*:

$$A \mapsto A'' = (A')' \text{ for } A \subseteq G \text{ (extent closure)} \quad (4)$$

$$B \mapsto B'' = (B')' \text{ for } B \subseteq M \text{ (intent closure)} \quad (5)$$

A concept (A, B) represents a maximal set of objects A that share a maximal set of attributes B . If another concept (A_1, B_1) is greater than (A, B) , it contains more objects but is described by fewer attributes, meaning that A_1 is more general than A . Thus, a subconcept-superconcept relation is defined:

$$(A, B) \leq (A_1, B_1) \Leftrightarrow A \subseteq A_1 \quad (6)$$

or equivalently,

$$(A, B) \leq (A_1, B_1) \Leftrightarrow B \supseteq B_1 \quad (7)$$

Given this relation, the formal concepts of a context generate a complete lattice, called *concept lattice* of the context. In this structure, every set of formal concepts has a *meet* (or greatest common subconcept), whose extent consists of those objects common to all extents of the set. Analogously, every set of formal concepts has a *join* (or least common superconcept), whose intent includes all attributes that all objects of that set of concepts have. The collections of all extents and all intents are intersection-closed.

The lattice can be graphically visualized as a *Hasse diagram* (Davey and Priestley, 2002), which may help understand the data. However, most of the times, lattices are extremely large, thus making visualization impossible. Figure 4 recaps all the main notions introduced in this section.

2.3 Linking KST and FCA

From the overviews that we saw, the reader may have already noticed a number of similarities between KST and FCA. However, not much work has been devoted to studying the two frameworks from a comprehensive perspective¹⁴.

Historically, Rusch and Wille, 1996 are the first to have pointed out interesting connections between the two theories. Specifically, they showed that the collection of the complements of the intents of a formal context is union-closed since the collection of the intents is intersection-closed (see Section 2.2). Therefore, the former can be seen as a knowledge space that is itself union-closed (see Section 2.1).

From this premise, they proposed a methodology to construct knowledge structures from formal contexts. As in FCA, the context is defined by the set G of individuals (treated as formal objects), the set M of questions (treated as formal attributes), and the incidence relation I , for which gIm expresses the fact that individual g correctly answered question m . In these terms, the intent of a formal concept is the set of questions correctly answered by a set of individuals and corresponds to a knowledge state.

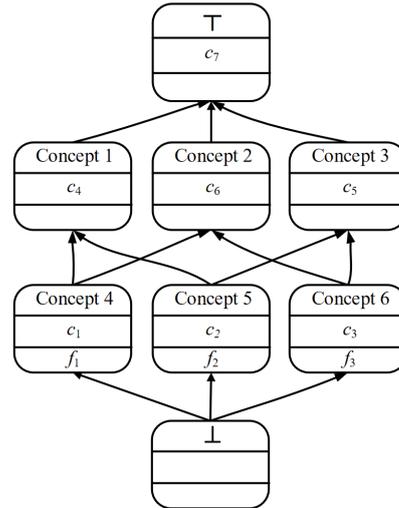
¹⁴See Nicotra and Spoto, 2019 for further reference.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
f_1	X			X		X	X
f_2		X		X	X		X
f_3			X		X	X	X

(A) FCA formal context.

\top	$\{f_1, f_2, f_3\}, \{c_7\}$
Concept 1	$\{f_1, f_2\}, \{c_4, c_7\}$
Concept 2	$\{f_1, f_3\}, \{c_6, c_7\}$
Concept 3	$\{f_2, f_3\}, \{c_5, c_7\}$
Concept 4	$\{f_1\}, \{c_1, c_4, c_6, c_7\}$
Concept 5	$\{f_2\}, \{c_2, c_4, c_5, c_7\}$
Concept 6	$\{f_3\}, \{c_3, c_5, c_6, c_7\}$
\perp	$\{\emptyset\}, \{c_1, c_2, c_3, c_4, c_5, c_6, c_7\}$

(B) Formal concepts of the formal context in Figure 4a.



(C) Hasse diagram representation of the concept lattice containing the concepts in Figure 4b.

FIGURE 4: Example of formal concept analysis taken from Chaikalis, 2016.

More recently, both Spoto, Stefanutti, and Vidotto, 2010 and Doignon, 2014 showed that the two theories share deeper connections. The formers demonstrated the possibility of building knowledge structures via the definition of skill maps, allowing to reconstruct a bijection between sets of skills (attributes) and sets of items (objects). The latter described that, being knowledge spaces particular partial ordered sets, they can be built using FCA following a workflow he presents.

In conclusion, from what we saw in this section, we can derive that the shared properties between FCA and KST can be exploited to build knowledge structures directly from formal contexts, representing relations between a set of formal objects (corresponding to KST's individuals) and a set of formal attributes (corresponding to KST's items).

2.4 Language Assessment Criteria

Language competence can be assessed using a wide range of exams and criteria, which can be cross-linguistic or language-specific.

The Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)¹⁵ is a widely accepted guideline used to describe linguistic skills and competences of students of foreign languages. It is used by teachers to create lesson plans, as well as by self-taught students to assess their skills independently, and by governments to redact national educational plans. Moreover, educational institutions use the CEFR scale when releasing language certificate results. It was created by the Council of Europe¹⁶ in the 1990s, with the main goal of making the general assessment of all European languages standardized. Although designed for

¹⁵<https://www.coe.int/en/web/common-european-framework-reference-languages/home>

¹⁶<https://www.coe.int/en/web/portal/home>

languages of Europe, it is now becoming increasingly widespread outside the EU as well (Little, 2007).

Its communication-oriented framework comprises six reference levels, namely, from the lowest to the highest, A1, A2, B1, B2, C1, C2. The skills and competences assessed range from vocabulary accuracy and grammatical competence to sociolinguistic and pragmatic use of the language. Both comprehension and production skills are included. See Figure 5 for a brief description¹⁷.

Proficient user	C2	Can understand virtually all types of texts. Can summarise information from different oral and written sources, reconstructing arguments and accounts in a coherent presentation. Can express themselves spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express themselves fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent user	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in their field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with users of the target language quite possible without imposing strain on either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise while travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
Basic user	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of their background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce themselves and others and can ask and answer questions about personal details such as where someone lives, people they know and things they have. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

FIGURE 5: CEFR levels' descriptors.

As far as Mandarin is concerned, the 汉语水平考试 *Hànyǔ Shuǐpíng Kǎoshì* 'Chinese Proficiency Test' (HSK)¹⁸ is the standardized test of language proficiency for non-native speakers. The test is administered by the Center for Language Education and Cooperation (formerly Hanban)¹⁹, an agency of the Chinese Ministry of Education. Its development began in 1984, and in 1992 it became a standardized test.

¹⁷A detailed and comprehensive explanation of all levels, skills, and competences can be found in Little, 2020.

¹⁸<http://www.chinesetest.cn/index.do>

¹⁹<http://www.chinese.cn/page/#/pcpage/mainpage>

Since its inception, HSK has been reshaped several times. The current format is organized on six levels, namely, from HSK 1 (the lowest) to HSK 6 (the highest), and tests reading, writing, and listening abilities²⁰.

This current format has been highly criticized by some scholars (Bellassen, 2011) for not being compliant with the CEFR levels and descriptors, despite Hanban's assertion that the six levels of the HSK correspond to the six levels of the CEFR. This is reflected in the process of reconciliation that the skills we crafted for this work had to undergo before being fully usable (see Section 3.2.1).

2.5 Related Work

The application of KST for language acquisition remains still challenging. However, this framework has successfully been employed to improve teaching and learning experiences of several other subjects. For instance, it has been applied to such subjects as mathematics (Ahlgren Reddy and Harper, 2013; Huang et al., 2016), chemistry (Taagepera and Arasasingham, 2013), stoichiometry (Arasasingham et al., 2004), and dance education (Yang et al., 2011).

ALEKS²¹ (Assessment and LEarning in Knowledge Spaces) (Falmagne et al., 2006) is a web-based intelligent learning and assessment system, the most well-known practical implementation of KST. The scholarly subjects offered by this tool are mathematics, chemistry, statistics, physics, and accounting. Practically, the website offers a probabilistic assessment test, aimed at uncovering the precise knowledge state of a pupil. After the results of the test, ALEKS presents a personalized learning plan, based on the skills the student is ready to acquire. Knowledge is re-assessed at regular intervals. The system has been available since the 1990s and adopted by many universities so far. Several analyses (Matayoshi and Cosyn, 2018; Cosyn et al., 2021) conducted on the efficacy of the tool have demonstrated that the use of ALEKS can aid teaching and learning at different levels of schooling. Its effectiveness has been shown to be at least comparable with traditional classroom teaching, if not better. Findings suggest as well slight positive impacts over speed and quality of learning and teaching.

Apart from ALEKS, other web-based tools have been developed as practical implementations of KST, such as RATH and KST Web. RATH (Hockemeyer, Held, and Albert, 1998) is a now defunct adaptive tutoring software applied to the field of elementary probability theory, which combined KST with a mathematical model for the structure of hypertext. Using implications about items, and the current knowledge state of the student, RATH pointed the individual only to documents containing concepts satisfying all the prerequisites. KST Web (Lee et al., 2014) was a free and publicly available web application for the construction and analysis of knowledge spaces that included some statistical metrics and the calculation of critical learning paths.

A specific application of the shared properties between KST and FCA can be found in Spoto, Stefanutti, and Vidotto, 2010. They used KST, jointly with FCA, for computerized psychological assessment. They developed a formal representation of the relations between the items of a questionnaire and a set of psychodiagnostic

²⁰To assess oral skills, another exam, the 汉语水平口语考试 *Hànyǔ Shuǐpíng Kǒuyǔ Kǎoshì* (HSKK), needs to be taken.

²¹<https://www.aleks.com/>

criteria. This representation is used to build a knowledge structure and develop an efficient adaptive tool for psychological assessment.

Despite the apparent lack of full-fledged applications of KST in CALL tools so far, other tutoring techniques have nonetheless been implemented with different success rates. However, most ITSs for CALL remain in the prototype phase or have been developed purely for research purposes (Slavuj, Kovačić, and Jugo, 2015; Tafazoli, María, and Abril, 2019). Nevertheless, some of these have been in use in real classroom settings. Among those, we can find E-Tutor (Heift, 2008; Heift, 2010) for German, TAGARELA (Amaral and Meurers, 2011) for Portuguese, and Robo-Sensei (Nagata, 2009) for Japanese.

To the best of our knowledge, the Revita project²², carried out by a group at the University of Helsinki, is the closest to our work. Revita (Katinskaia, Nouri, and Yangarber, 2018) is a language learning web-based platform²³, which can be placed at the intersection of CALL and ITS. They claim to be “currently the only platform for learning/tutoring beyond the beginner level, that is functional, freely-available and supports multiple languages”. This tool tries to address both CALL and ITS requirements, among which we can find automatic assessment of students’ knowledge and adaptive learning (see Section 1.1).

In Katinskaia, Nouri, and Yangarber, 2018, the authors proposed implementing KST in Revita to address some of the issues mentioned above but left it for future work. Steps towards its implementation have been taken in subsequent works. In Hou et al., 2019, the authors made use of a modified Elo formula (Elo, 1978) to obtain ratings for learners’ proficiency and the difficulty of the linguistic concepts that the learners are trying to master.

The Elo rating was originally used for rating the skills of chess players. However, modified versions of it have been used in a variety of domains. Essentially, the Elo formula defines the expected result of player A in a match against player B . The idea put forward in the paper is to simulate “battles” between students and concepts to master. The modified version of the Elo formula used in the paper has the following form:

$$E_A = \frac{1}{k} \cdot 1 + \left(1 - \frac{1}{k}\right) \cdot \frac{1}{1 + 10^{\frac{R_C - R_S}{\sigma}}} \quad (8)$$

where k is the number of choices in the multiple-choice questions presented to the students, R_C the current rating of concept C , R_S the current rating of student S , and σ a scale variable set to 600.

Further steps to implement KST to Revita were taken in Vu, 2020. In this work, the author adapted BKT (see Section 1.1) and grow-shrink Markov blanket algorithm (Margaritis, 2003) to craft knowledge states and track student’s advancement in Russian language learning. He eventually succeeded in modeling a network structure in the form of a surmise graph (see Section 2.1) of the knowledge domain.

²²<https://www2.helsinki.fi/en/projects/revita-language-learning-and-ai>

²³<https://revita.cs.helsinki.fi/home>

Chapter 3

Implementation

This chapter is devoted to the presentation of the methodology and the description of the implementation steps that we followed for the realization of the work. We will open the chapter with Section 3.1 outlining the methodology adopted. Then, we will start detailing the implementation's steps. In Section 3.2, we will present the data acquisition campaign launched to gather information from learners. After that, in Section 3.3, we will describe how we developed the tutoring system and the functionalities made available. Specifically, Section 3.3.1 will be dedicated to the process followed to construct the learning space starting from a formal context, Section 3.3.2 to the description of the assessment process we developed, while Section 3.3.3 to the algorithm we proposed for the calculation of the best learning paths. Finally, in Section 3.4, we will illustrate the integration process of the tutoring system in the ecosystem of Naimrod.

3.1 Methodology

First and foremost, a quick note on the terminology that will be adopted throughout the rest of the thesis. In the previous chapters, we saw that the two mathematical frameworks share several properties but use different technical terms to indicate similar notions. By virtue of the similarities between the formalisms, we would like to reconcile the technical terms of the two mathematical formalisms. To do so, we decided that, for the rest of the thesis, we will employ interchangeably²⁴:

- *skill, item, attribute* to indicate a discrete notion that a student should master;
- *question, instance* to indicate a specific, concrete example of a discrete notion;
- *intent (of a formal concept), knowledge state* to indicate a specific set of discrete notions that an individual may know at a specific time;
- *individual, object, student, subject, test-taker* to indicate a human entity who undertakes tests provided by the tool.

Before delving into the project's implementation details, we will quickly sketch the methodology adopted to develop the work. The core idea is to leverage FCA's versatility and the shared mathematical properties between KST and FCA as a basis to construct a knowledge structure.

²⁴Eventually, if the need to discern the different terms arises, we will highlight it on a case-by-case basis.

The first step is to build a relevant knowledge domain representative of the steps needed to reach full proficiency in Mandarin. Next is to acquire pertinent data containing students' response patterns to the set of items identified.

Once the data are gathered and shaped into a binary table, the formal context is created and the formal concepts computed. The shared properties between FCA and KST now enter the equation: the intents of the formal concepts correspond to the states of the knowledge structure. Hence, at this moment, the underlying formal structure is ready, and the assessment process and the algorithm for calculating the best learning paths can be built on top of it. Both algorithms depend on a dynamic exploration of the mathematical structure without the need to compute the whole lattice at once beforehand.

The assessment process uncovers the exact knowledge of an individual through the computation of adjacent knowledge states, updating at each response the current knowledge state of the individual.

The algorithm for computing the best learning paths has two modes. The first one, interactive, allows the user to select the next state at each step, while the second one, based on a scoring function, automatically selects the best learning path from a starting point to an endpoint.

3.2 Data Acquisition

The launch of a data acquisition campaign was motivated by the lack of relevant data for our purposes. Our tool aims to construct a learning space starting from an FCA's formal context. Thus, we needed to extract binary tables from students' response patterns to a set of problems.

However, despite the existence of learners' corpora for Mandarin (Lu and Chen, 2019), their availability is limited. Moreover, finding or adapting resources specifically to our objective is difficult. Indeed, none of the corpora we tested presented the characteristics that we needed to develop the system. In particular, extracting response patterns from those corpora seemed hardly feasible. Consequently, we decided to create an online platform to assess students and gather binary data ourselves. This campaign required collaboration with a Mandarin teacher working for the French Ministry of Education.

3.2.1 Selection of the Skills

The first step in the data acquisition campaign was to choose the set of problems, namely, the spectrum of skills, on which we would assess learners. Options included broad HSK and CEFR levels (see Section 2.4) and fine-grained lexicogrammatical skills (such as "Can distinguish between different meanings of x "), among others.

Eventually, we deemed it necessary to reach a compromise in skills' granularity without hampering the quality of the system. Had the number of skills been too large, the amount of effort required for crafting skills and posing questions to students would have been unsustainable in such a limited amount of time. On the other hand, had the number been too small, we might have had problems achieving a significant ordered structure to analyze.

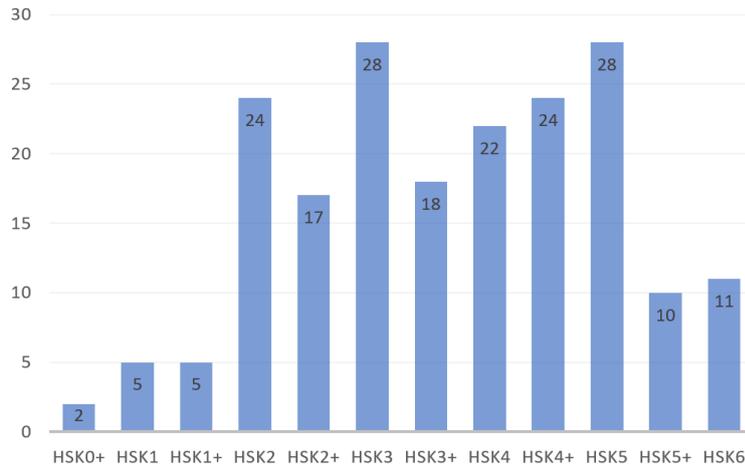


FIGURE 6: Distribution of the 195 skills with respect to their HSK tags.

Thus, we decided to select representative HSK²⁵ and CEFR²⁶ descriptors to have comprehensive coverage in linguistic abilities to assess. For this campaign, we limited ourselves to listening, reading, and audio-visual comprehension since acquiring these data automatically is a more straightforward task. Future work includes broadening the range of skills, including production skills.

The selected skills needed to be modified before being entirely usable. This reconciliation process included adapting existing skills, adding missing ones, and tagging each of them with appropriate HSK and CEFR levels. Eventually, we ended up having 195 skills²⁷, covering phonological, orthographic, syntactic, lexical, logical, sociolinguistic, and pragmatic competences. The full list of skills can be found in Appendix A. Refer to the bar chart in Figure 6 for the distribution of the skills with respect to the HSK tags.

The following step was to associate each skill to a number of concrete examples that would be posed to the test-takers to assess their mastery of each skill. Following KST’s item-instances dichotomy (see Section 2.1), we paid particular attention to craft, for each skill (item), three questions (instances) that would have the same difficulty and cover the same content. The choice of creating three questions for each skill was made in order to mitigate the possibility of lucky guesses. Ideally, we would consider a subject to have achieved mastery on a determined skill if and only if they responded correctly to at least two out of three questions.

The questions were created from scratch based on native content and other online resources or taken from past HSK exams. We deemed it essential to diversify both the content and the type of questions to improve the user experience and, most importantly, minimize the chances of careless errors due to boredom. Thus, we created several types of exercises, such as multiple-choice questions and rearrange-the-sentence, using different media such as texts, audios, videos, and pictures.

²⁵http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/202103/W020210329527301787356.pdf

²⁶<https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>

²⁷This number was supported by the fact that the order of magnitude in previous successful applications, such as ALEKS, is precisely in the order of hundreds.

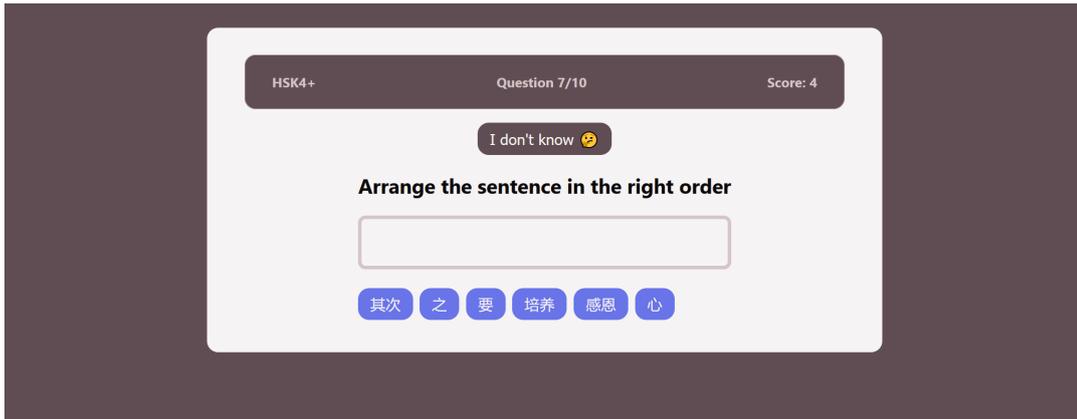


FIGURE 7: WenTest’s interface.

3.2.2 WenTest

WenTest²⁸ is the online platform that we built to host questions created following the methodology mentioned in Section 3.2.1 and gather data from individuals. We built it in such a way as to provide test-takers with the smoothest and most engaging experience possible. We opted for a minimalistic and friendly user interface and included elements of gamification (such as a global ranking), following as well feedback from our beta-testers. A screenshot of the interface can be seen in Figure 7.

Users participating in this campaign created a profile (after accepting a GDPR agreement) and were prompted with 585 questions²⁹. To offer test-takers a more pleasant experience and, most importantly, avoid careless errors, we split the test into batches of 10 questions and allowed users to progress through the batches across separate user sessions.

Moreover, in order to guarantee fairness in the responses, we took measures to prevent subjects from cheating. In particular, we do not allow copy-pasting the text of the questions.

When users take the test, we keep track of the time that elapses between when they are prompted with a question and when they give the answer. This provides valuable information regarding the difficulty of each question. Once sufficient data will be available, Naimrod will exploit this information to refine the scoring of the skills (see Section 3.3). Furthermore, we encourage individuals to skip questions they do not know (instead of giving a random answer, for example), as it has been shown to boost the performances of learning spaces (Cosyn et al., 2021).

In typical applications of KST, much effort is put into avoiding potential lucky guesses. Accordingly, multiple-choice questions are rarely used. However, in our application, we needed to use multiple-choice questions for rapid assessment and quick parsing of the answers. We, nonetheless, took care of presenting a large and consistent number of possible responses to mitigate lucky guesses as much as possible.

An analysis of the data we managed to gather so far using this platform and the description of the experiments conducted with them can be found in Section 4.2.

²⁸<https://wentest.ovh.naimrod.com/>

²⁹As a reminder, we crafted three questions for each of the 195 skills to assess, thus giving a total of 585 questions; see Section 3.2.1.

3.3 The Learning Space

The data we gathered from WenTest served as the basis for the construction of the learning space. The data were restructured into a binary table T , corresponding to a formal context, which is the starting point of our methodology. Each row corresponds to the response pattern v_j obtained by an individual j , while each column to one of the 195 skills. Each cell T_{ij} corresponds to an incidence relation between individual j and skill i . If individual j has mastered skill i , then a 1 occupies the cell, otherwise a 0. Refer to Table 2 for an extract of the dataset.

	skill 1	skill 2	...	skill 194	skill 195
individual 1	1	0	...	0	0
individual 2	1	1	...	0	0
...
individual j	1	1	...	1	1

TABLE 2: Sample dataset in the form of formal context.

Each skill was given a *difficulty score* based on two factors: (1) results from IRT's graded response model (GRM) (Samejima, 1997) applied to the dataset, and (2) HSK and CEFR tags of the skills³⁰. The use of multiple factors was motivated by the fact that we wanted to achieve scores that would be as precise as possible with respect to the actual difficulty of the skill.

In particular, the bulk of the difficulty scores' assignment comes from GRM. GRM is a mathematical model belonging to IRT (see Section 1.1) used to evaluate students' performances from their response patterns to a test and rank the difficulty of the different abilities assessed in the test. It can deal with both dichotomous and polytomous categories. Specifically, for our application, we used a Python implementation of dichotomous GRM³¹ that uses marginal maximum likelihood (MML) (Bock and Aitkin, 1981) to estimate the difficulty of each skill. In detail, the formula for the MML version of GRM is the following:

$$L(a_i, b_{u_i}) = \prod_{j=1}^N \int_{-\infty}^{\infty} g(\theta_j) P_{v_j}(\theta_j) d\theta_j = \prod_{j=1}^N P_{v_j} = \prod_V P_v^{r_v} \quad (9)$$

where a_i and b_{u_i} are the parameters to estimate, N is the number of individuals in the dataset, θ_j is the ability of an individual j , $g(\theta_j)$ is the ability density function, v_j is the response pattern obtained by an individual j , P_{v_j} is the marginal probability for the response pattern v_j of an individual j , r_v is the frequency of a specific response pattern v , and \prod_V indicates the product over all possible response patterns. See Samejima, 1997 for a more detailed explanation.

The results given by this model are then combined with the HSK and CEFR tags assigned to each skill during the reconciliation process mentioned in Section 3.2.1.

The difficulty scores obtained by the combination of the above factors are utilized as a scoring function for conveniently ordering the skills in the assessment process (see Section 3.3.2) and automatically selecting the best next step in the construction of the learning path (see Section 3.3.3).

³⁰In the future, information on the time needed to respond to the questions on WenTest will be included.

³¹<https://github.com/eribeau/girth>

3.3.1 Space Construction

The construction of the mathematical structure required much effort. At first, we tried to construct a complete concept lattice since several off-the-shelf tools can be found online to this avail³². However, it should be borne in mind that the long-term objective of this software is to give the possibility to teachers to assess pupils on hundreds, if not thousands, of skills. Consequently, the formal context that should be handled may get very large. Thus, we experimented the feasibility of using the tools mentioned above in such a computationally expensive application. In order to do so, we created artificial formal contexts with hundreds of students and skills. This way, we mimicked potential full-fledged applications of this software at a large scale.

Conducting the above experiments, we realized that all the tools, even those that implement the most efficient algorithms known (Strok and Neznanov, 2010), incur memory errors or run for hours without giving results when prompted with large contexts (see Section 4.1 for an analysis on this matter). Hence, we discovered that using off-the-shelf tools for our application is practicable neither under time nor space complexity. This led us to discard a global approach based on the construction of a complete lattice.

In order to achieve a stable software architecture, which would not be limited by the order of magnitude of skills and students, we thus opted for a lightweight and dynamic local approach that would allow us to build and explore only the parts of the space needed by each user's request. This way, what needs to be stored is no more than the formal concepts with their intents and extents.

At this stage, once all concepts together with intent-extent pairs are obtained, the shared mathematical properties between FCA and KST (see Section 2.3) enter the picture. The intents of the formal concepts just calculated correspond precisely to the knowledge states of the knowledge structure we are aiming at obtaining. Thus, the structure can now be dynamically explored.

The exploration, which is the focal point for assessment and learning path algorithms (respectively detailed in Section 3.3.2 and Section 3.3.3), is based on the search of adjacent states and their fringes, traversing only the part of the space we are interested in at each request.

3.3.2 Assessment Process

The fundamental starting point in our tutoring system is the assessment process. Through a limited number of questions posed to the student, this process aims to uncover the exact knowledge they have acquired. Based on the results of this process, the system proposes the skills that the student is probably most ready to acquire next.

The assessment process that we developed for our learning space is loosely inspired by Chen, 2013. In the paper, the author introduced a procedure to dynamically compute the fringes of a knowledge state according to prerequisite relations in the surmise graph, on top of which he developed an adaptive assessment process using inner and outer fringes of knowledge states.

The algorithm described in this section is sketched in Algorithm 1. Our adaptive assessment process is meant to be interactive. At each step of the process, an instance of the skill to be assessed is shown to the student (*instance()* in Algorithm 1).

³²<https://upriss.github.io/fca/fcasoftware.html>

Algorithm 1: Algorithm for dynamic assessment process.

Result: Accurate knowledge state of the student

- 1 SD = set of all skills in the domain ordered by discrimination factor;
- 2 SI = set of all intents;
- 3 **begin**
- 4 $response = False$;
- 5 **while** *not* $response$ **do**
- 6 $s \leftarrow BinarySearch(SD)$;
- 7 $response \leftarrow instance(s)$;
- 8 **if** $response$ **then**
- 9 $introSkill \leftarrow s$;
- 10 **else**
- 11 $SD \leftarrow SD \setminus \{s\} \cup surmisers(s)$;
- 12 $I \leftarrow \{introSkill\}''$;
- 13 $currentKS \leftarrow I$;
- 14 $N \leftarrow \min(\{currentKS \cup \{m\}'' | m \in M\})$;
- 15 $S \leftarrow \cup\{n \setminus currentKS | n \in N\}$;
- 16 **while** $S \neq \emptyset$ **do**
- 17 $i \leftarrow median(ordered(S))$;
- 18 $response \leftarrow instance(i)$;
- 19 **if** $response$ **then**
- 20 $currentKS \leftarrow currentKS \cup \{i\} \cup surmised(i)$;
- 21 $N \leftarrow \min(\{currentKS \cup \{m\}'' | m \in M\})$;
- 22 $S \leftarrow \cup\{n \setminus currentKS | n \in N\}$;
- 23 **else**
- 24 $S \leftarrow S \setminus \{i\} \cup surmisers(i)$;
- 25 $initialKS \leftarrow currentKS$;
- 26 **return** $initialKS$

First, the process needs to identify an introductory skill mastered by the student. To uncover this first skill, the algorithm starts by showing instances of the most discriminant skills³³ until the student responds correctly to their first question. In this first part of the process, we perform a binary search (Hall and Bellman, 1960) over the most discriminant skills to speed up the search.

Once the introductory skill has been identified³⁴, we compute the smallest intent containing this skill. We consider this intent to be the introductory knowledge state of the subject, which will be refined in the following steps of the assessment process.

After that the introductory state I has been determined, the algorithm finds its upper adjacent states and retrieves the skills available for assessment, namely the set S of all the skills such that adding one or more skills to I yields an adjacent state. It is important to note that we cannot talk of outer fringe in terms of how we defined it in Section 2.1 since we cannot assure that an adjacent state is yielded by adding precisely one skill to I . Indeed, it may be the case that two or more skills need to

³³The most discriminant skills are considered to be those mastered by at least 50% of the whole population present in the original dataset.

³⁴If none of the skills has been identified to be the introductory one, we consider the student to be an absolute beginner with no prior knowledge of the subject.

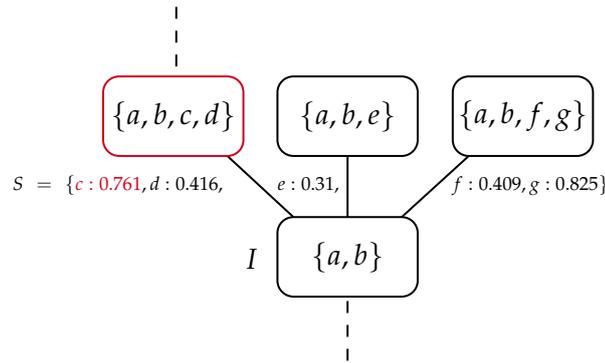


FIGURE 8: Visualization of a step of the assessment process: I is the introductory state, c in red is the skill with the median score in S , and the state highlighted in red is the updated knowledge state of the student after a correct response to an instance of c .

be added to I to yield an adjacent state. This is good during the assessment process since we actually want to exploit implications to speed up the search³⁵.

At this stage, the skills in S are ordered by difficulty based on the scoring function we described in Section 3.3 (*ordered()* in Algorithm 1). An instance of the median skill i (*median()* in Algorithm 1) of the collection S is shown to the subject. If the response given by the subject is correct, we update the current knowledge state of the subject, adding i and all the skills prerequisite to i (*surmised()* in Algorithm 1). After that, S is recomputed based on the adjacent states of the updated knowledge state of the subject. Otherwise, if the response given is wrong, we remain in the same knowledge state and remove i from the domain of available knowledge together with all the skills having i as a prerequisite (*surmisers()* in Algorithm 1). The algorithm keeps showing an instance of the median skill of S at each step until S gets emptied. The knowledge state at the last step of the algorithm is considered to be the actual set of skills mastered by the subject.

To clarify, Figure 8 presents a concrete example of the aforementioned process. The introductory knowledge state I contains the items $\{a, b\}$. After its adjacent states $\{a, b, c, d\}$, $\{a, b, e\}$, and $\{a, b, f, g\}$ are found, the algorithm retrieves the set S of skills available for assessment: $\{c, d, e, f, g\}$. An instance of c , the skill with the median score among the skills in S , is shown to the student. In this example, we assume that the student responds correctly to the instance shown. Therefore, we update the knowledge state of the student, adding c and its prerequisite d .

As a final note, the assessment process clearly cannot be used in the absence of students. Nonetheless, we allow users to set a customized starting point by manually selecting an initial knowledge state. They can do that by setting any HSK or CEFR level as the starting point or manually selecting the skills to be in the initial knowledge state. In the former case, all the skills up to the level selected are considered the initial knowledge state. In the latter case, the initial knowledge state is the smallest state containing at least all the skills included in that customized selection.

³⁵On the contrary, when calculating the best learning path, we need to guarantee a granular transition from any state to its adjacents in order to satisfy learning smoothness and learning consistency. Refer to Section 3.3.3 for the solution we conceived.

Algorithm 2: Algorithm for calculating the optimal learning path.

Input: Initial knowledge state IKS , objective knowledge state OKS

Output: Best learning path $Path$

```

1 begin
2    $Path = \{\}$ ;
3    $currentKS = IKS$ ;
4   while  $currentKS \neq OKS$  do
5      $N \leftarrow \min(\{currentKS \cup \{m\}'' | m \in M\})$ ;
6     if  $any(\{|n \setminus currentKS| | n \in N\}) > 1$  then
7       forall  $n | |n \setminus currentKS| > 1$  do
8          $N \leftarrow N \setminus n$ ;
9          $X \leftarrow X \cup \{n \setminus currentKS\}$ ;
10     $N \leftarrow N \cup \{currentKS \cup \{x\} | x \in X\}$ ;
11     $OuterFringe \leftarrow \bigcup \{n \setminus currentKS | n \in N\}$ ;
12     $bestSkill \leftarrow best(OuterFringe)$ ;
13     $Path \leftarrow Path \cup \{bestSkill\}$ ;
14     $currentKS \leftarrow currentKS \cup \{bestSkill\}$ ;
15 return  $Path$ 

```

3.3.3 Learning Path

In the previous section, we saw that the assessment process allows determining the current knowledge state of the student. This knowledge state is the starting point for the calculation of the best learning paths.

Another fundamental condition that needs to be set for this calculation to work is selecting the final knowledge state, namely the pupil's desired learning objective. This final objective can be any HSK or CEFR level. Technically, the final knowledge state is considered the set containing all the skills up to the level chosen as the final objective, plus eventual additional skills already mastered by the student, which are beyond the selected objective.

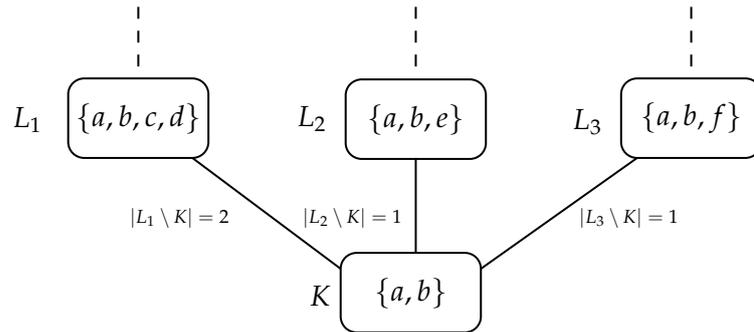
The aim of this process is to find the best possible ordered sequence of knowledge states given the initial state of the student and the final objective state of learning. The algorithm is outlined in Algorithm 2.

Our algorithm starts by finding the upper adjacent states of the initial knowledge state. If we name K the current knowledge state, for each adjacent state L_i of K , we check $|L_i \setminus K|$. If any $|L_i \setminus K| > 1$, learning smoothness and learning consistency (the two additional conditions that make a knowledge space a learning space; see Section 2.1) are not satisfied.

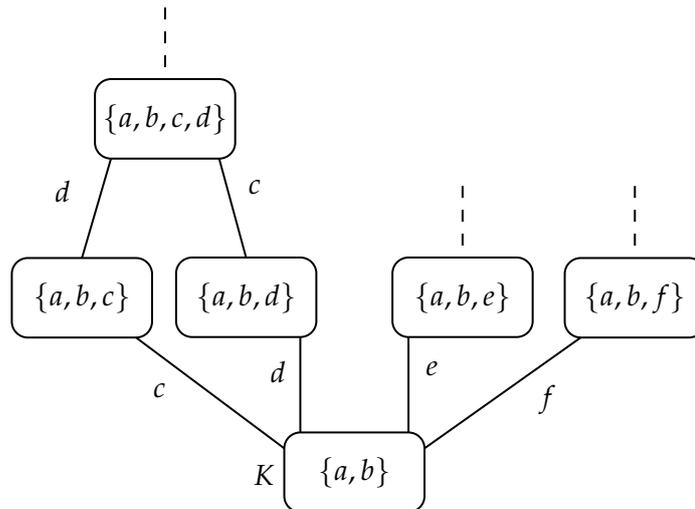
Therefore, we devised a naive but effective solution at this stage of the algorithm to meet these two conditions: for each $|L_i \setminus K| > 1$, we compute the powerset $\mathcal{P}(L_i \setminus K)$. This way, for each $|L_i \setminus K| > 1$, we retrieve all the possible learning combinations of the items in $L_i \setminus K$ and achieve learning smoothness and learning consistency.

Once the two conditions above are met, the proper outer fringe of the current knowledge state is obtained, and the search for the best learning path can progress.

The following step of the algorithm is the selection of the best skill among the current outer fringe. To do so, we crafted two different modalities: interactive and



(A) A section of the space when learning smoothness and learning consistency are not met.



(B) The same section of the space after the computation of the powerset $\mathcal{P}(L_1 \setminus K)$.

FIGURE 9: Visualization of a step of the algorithm for the selection of the best learning path.

automatic. The interactive mode allows the user to manually pick the skill they believe the student to be more ready to acquire. Conversely, the automatic mode selects the most accessible skill in the outer fringe, based on the scoring function we described in Section 3.3. When the skill is selected, it is added to the learning sequence, and we move up in the space: the current knowledge state becomes the adjacent state containing the skills at step $t - 1$ plus the best skill selected at the current step t .

The algorithm continues until the set of current skills is equal to the set of final skills, meaning that the algorithm reached the knowledge state that the user initially set as an objective, and the optimal path is found. Most importantly, the skills are now ordered in the best possible learning sequence.

A concrete example of one step the algorithm explained above can be found in Figure 9. Considering $K = \{a, b\}$ as our initial state, the algorithm starts by finding the upper adjacent states of K in $L_1 = \{a, b, c, d\}$, $L_2 = \{a, b, e\}$, and $L_3 = \{a, b, f\}$. Then, we check the cardinalities $|L_1 \setminus K|$, $|L_2 \setminus K|$, and $|L_3 \setminus K|$, and we find that $|L_1 \setminus K| > 1$. This means that learning smoothness and learning consistency are not met. We can see the status of the space at this point of the algorithm in Figure 9a. To

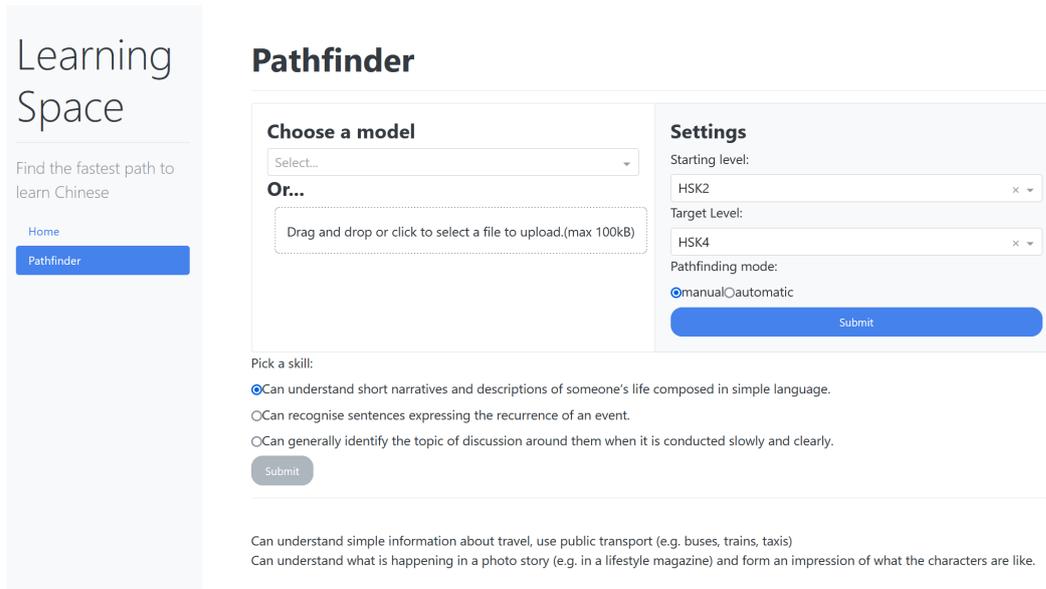


FIGURE 10: Learning Space’s interface.

satisfy the two conditions that would let us obtain a learning space, we need to achieve a one-by-one transition from K to L_1 . Thus, we calculate $\mathcal{P}(L_1 \setminus K)$ and retrieve all the possible learning combinations for $\{c, d\}$: c, d and d, c . Figure 9b shows what the space looks like after this computation. After achieving local learning smoothness and learning consistency, the outer fringe of K can be easily retrieved: $\{c, d, e, f\}$. At this stage, we need to pick the best skill among the outer fringe. Let us consider that this best skill is c , then c is added to the best learning sequence and we move to state $\{a, b, c\}$.

3.4 Integration into Naimrod’s Framework

Reiterating what was discussed in Section 1.2, the objective of building the learning space for Mandarin acquisition is to improve Naimrod’s service, introducing an ITS in a website that offers CALL. This section deals with the aspects that have deserved consideration when discussing the integration of the product of research into Naimrod’s societal framework.

First of all, the commercial implications for Naimrod deriving from integrating the learning space may be considerable. Given that the final aim of their service is to reduce the time that teachers spend crafting engaging and tailor-made lessons for pupils, a tool that is flexible enough to aid teachers in dynamically ordering the skills that a student is most probably ready to acquire would certainly be highly valued. Such a feature would let teachers save time and focus on creating involving lessons. Moreover, the tool could be employed for interactive assessment of students, suggesting to teachers the best skills to assess based on previous answers, speeding up tests.

The second aspect is the technical execution of the integration. The idea is to implement this work in an interface where teachers would have all the features at their disposal. Teachers would first have the possibility of assessing pupils using the assessment process and retrieve their actual knowledge states and adapt their lesson

plans accordingly. At the time of writing, the tool is in the prototype phase³⁶. Future plans include making it available as a service. A screenshot of the tool is presented in Figure 10.

As for the back-end, the method we adopted is computationally not demanding and versatile. As we have seen, the tool would be able to handle contexts containing thousands of skills and students. Moreover, the experiments we conducted (see Sections 4.3 and 4.5) indicate that the tool may be versatile enough to adapt to other languages. It would need a sufficient amount of data to output an exploitable ordered structure. This is a favorable outcome since Naimrod envisions the enlargement of the pool of covered languages in the near future.

Finally, the last aspect to consider is the target audience. Currently, the tool has been developed to meet the needs and requirements of teachers. We have deemed it crucial to plan the creation of an easy-to-use tool since many teachers may be technophobic and unwilling to learn how to use complex software. Eventually, plans are to enlarge the target audience and make the tool available to students, as well.

³⁶<https://lsgui.ovh.naimrod.com/>

Chapter 4

Experiments and Results

The aim of this chapter is to provide a description of the experiments we performed to build and evaluate our tool. Throughout the process, we devised experiments trying to address mainly the following questions:

1. Can KST be applied to language acquisition and teaching?
2. Is a learning space for languages useful to teachers and students?
3. Can we find implications between skills, which would help uncover different learning strategies?

The rest of this chapter will try to give some answers to these questions. In particular, in Section 4.1, we will present an analysis regarding the issues surrounding the construction of the learning space. Then, in Section 4.2, we will detail the experiments we conducted using WenTest’s data. In Section 4.3, we will introduce a new dataset consisting of response patterns coming from Russian learners and present the results obtained working with it. Next, in Section 4.4, we will present the opinions of the teachers that emerged when they used the first version of our system. Finally, we will conclude the chapter with Section 4.5 addressing in a closing discussion the questions posed above.

As a general remark, it should be underlined that the results presented in the following sections are meant to be neither final nor comprehensive. What we present here are preliminary results that we could achieve with limited time and resources.

4.1 Lattice Construction

As mentioned in Section 3.3.1, we approached the issue of constructing the space under different perspectives, each with its pros and cons. In this section, we will detail the experiments we conducted.

As a first choice, we tried to see if building a complete lattice was something possible. Achieving it would mean having the full knowledge structure at disposal without further processing. To test the feasibility of this approach, we tried out several off-the-shelf tools available online. Experiments were conducted on a laptop mounting an Intel Core i7 processor at 1.80 GHz using 8 GB of RAM, running Ubuntu version 20.04.

The tools tested and the results obtained can be seen in Table 3. Each tool was input with artificially created formal contexts of different dimensions. From the table, we can see that almost all the tools succeeded in building a partial order when input with small-sized contexts. When moving towards mid-sized contexts, all the

	100 × 100	500 × 500	1,000 × 1,000	5,000 × 5,000	10,000 × 10,000
ToscanaJ	✓	✓	✗ T	✗ T	✗ T
ConExp	✓	✓	✗ T	✗ T	✗ T
Galicia	✓	✓	✗ T	✗ T	✗ T
GALACTIC	Unable to install				
FCALGS	✓	✓	✓	✗ M	✗ M
In-Close	✓	✓	✓	✗ M	✗ M
kst	✗ T	✗ T	✗ T	✗ T	✗ T

TABLE 3: Experiments with several off-the-shelf tools. Columns values indicate the dimension of the input formal context. T = time error. M = memory error.

tools (except FCALGS and In-Close, which are the ones that run the most optimized algorithms) started incurring time errors, running indefinitely. Finally, when the tools were fed with big-sized formal contexts, even these two tools eventually run into memory errors, halting the machine.

Thus, these experiments showed that a global approach based on the construction of complete lattices is not scalable since any tool will have problems when prompted with large inputs. Running the experiments on a more powerful machine will undoubtedly improve the results. However, we did not stop our research and kept looking for something that would not be constrained by the size of the context or the power of the machine.

Another approach we attempted was building the partially ordered set of the attribute and object concepts (AOC-poset) (Osswald and Petersen, 2003; Petersen, 2004). The *AOC-poset* is a simplified version of a concept lattice. In particular, it is a subset of the concept lattice that considers only concepts introducing an object or an attribute. AOC-posets are much smaller and easier to compute than concept lattices, given their reduced combinatorial power.

Specifically, for our experiments, we tried *AOC-poset Builder*³⁷ (Dolques, Le Ber, and Huchard, 2013). Using this tool, we successfully built structures starting from huge formal contexts (10,000 × 10,000). However, an AOC-poset is not useful in a learning space setting since the partial order returned is too simplistic. In detail, the AOC-poset is not as informative as the concept lattice, representing only the concepts that introduce an object or an attribute. Therefore, since it eliminates that granularity fundamental in learning spaces, it is not applicable in this setting.

Eventually, to achieve a stable software architecture, we opted for an approach based on a local exploration of the space without the need to construct it beforehand. This method is based on the only computation of the formal concepts with their intents and extents without the relations of adjacency between them. This crucial information is uncovered only later during each user’s request and only on the concepts needed at each search. A Python package, *Concepts*³⁸, helped us compute the formal concepts from the formal context. This method is extremely scalable since computing concepts is not demanding and can handle very large datasets.

Indeed, computing the formal concepts of a 10,000 × 10,000 context took a few seconds. Finding the best path is also quick: the search of the adjacent states is almost

³⁷<http://www.lirmm.fr/AOC-poset-Builder/>

³⁸<https://github.com/xflr6/concepts>

instantaneous, while computing a full path on the data gathered from WenTest (see Section 4.2.1) took 56 seconds.

4.2 Mandarin Data

Gathering data from WenTest (see Section 3.2.2) proved more challenging than expected. First, crafting and correcting skills and instances were demanding tasks. Second, the development of WenTest was delayed by unforeseen factors³⁹. These, together with the fact that taking the test itself is enormously time-consuming, are the reasons why the data collection campaign did not turn out to be as profitable as hoped.

We eventually acquired exploitable response patterns by seven individuals. To achieve accurate predictions, a learning space would need thousands of response patterns. Clearly, gathering successfully such a massive amount of data was out of the scope of the internship. Plans are to keep gathering data and incrementally inject them into the space to improve and refine the predictions. Anyway, we tried to do experiments with the data we had at our disposal. Surprisingly, the preliminary results that we obtained are encouraging.

4.2.1 Analysis of the Structure

Using the response patterns acquired from WenTest, we computed the best learning path from the empty state to the full state using the automatic mode (see Section 3.3.3). We succeeded in visualizing the concept lattice, presented in Figure 11. The best learning path is highlighted in red, while the corresponding learning sequence can be found in Table 5 in Appendix B. Technically, the lattice in Figure 11 is a knowledge structure since it does not satisfy the learning smoothness⁴⁰ required to qualify it as a learning space.

This lattice should be interpreted side by side with the sequence in Table 5. The skills within a subsection of the table are the ones that are acquired in the corresponding position in the lattice. To ensure a gradual acquisition of all the skills, learning smoothness is achieved by following the algorithm presented in Section 3.3.3.

Several deductions can be made by analyzing the structure and the path. Among the positive points, we can notice that the shape of our structure indeed reflects the classical shape of a knowledge structure, where the concentration of knowledge states at the bottom of the structure is lower than in the middle or at the top of it. Consequently, more implications between items can be drawn.

Studying the learning sequence, we can observe that in some sections of it, the advancement seems to be following the natural progression of students when learning Mandarin. For instance, the sequence tells us that a plausible progression is the following:

1. “*Can recognise words/signs, names and numbers that they already know in simple, short recordings, provided these are delivered very slowly and clearly.*” (acquired in *Concept 1*)

³⁹For more details regarding these issues, refer to Section 5.1.

⁴⁰Learning smoothness is however achieved during the exploration of the portions of the space needed to find the desired learning path.

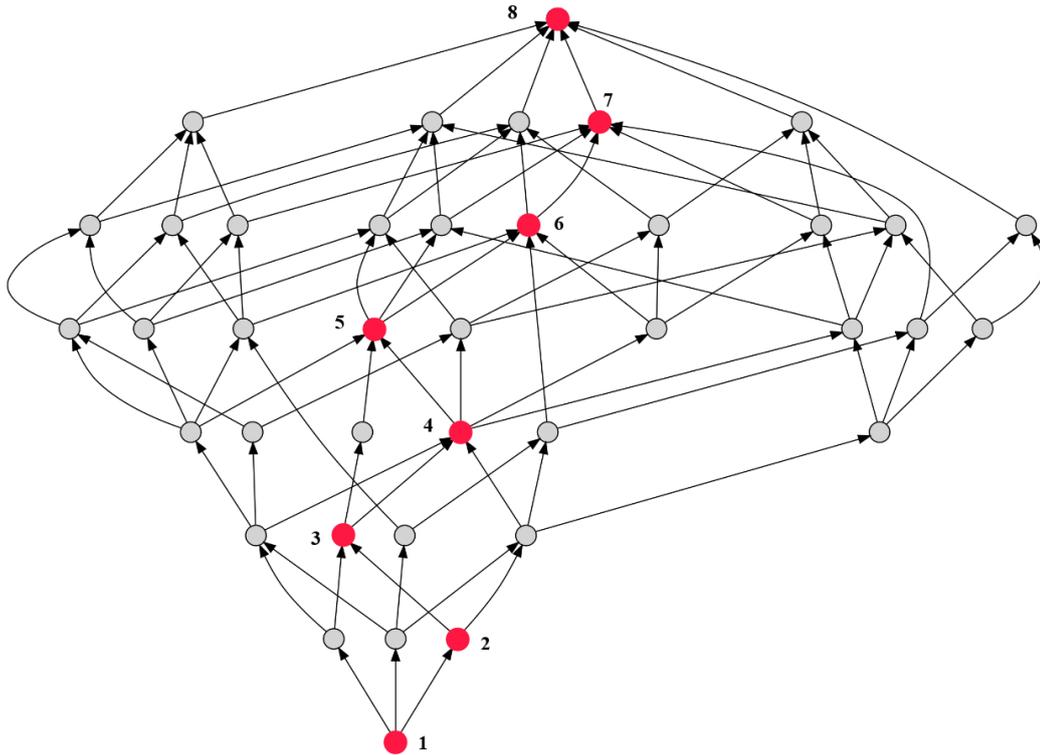


FIGURE 11: Knowledge structure built using the data gathered from WenTest, with the best learning path highlighted in red.

2. “Can understand short, simple dialogues between two persons provided they are delivered slowly and clearly in a defined, familiar everyday context.” (acquired in Concept 4)
3. “Can understand the main points of what is said in a straightforward monologue (e.g. a guided tour), provided the delivery is clear and relatively slow.” (acquired early in Concept 5)
4. “Can understand vlog-type content about familiar subjects delivered at a normal speed with cuts.” (acquired late in Concept 5)

This progression is indeed reasonable since a student first catches isolated words and expressions in familiar contexts and only later understands monologues and vlogs. However, some advancements proposed by the sequence are extremely inaccurate. For example, the sequence suggests a progression like the following:

1. “Can recognize and find specific information on a document that includes culture-specific content, like a restaurant menu.” (acquired early in Concept 4)
2. “Can recognise times and places in very simple notes and recordings from friends or colleagues (e.g. “Back at 4 o’clock” or “In the meeting room”).” (acquired late in Concept 4)
3. “Understands basic ways to link parts of sentences.” (acquired in Concept 5)

4. “Can recognise basic Chinese radicals⁴¹ in characters.” (acquired in Concept 7)
5. “Can understand everyday signs and notices, etc. in public places, such as streets, restaurants, railway stations; in workplaces, such as directions, instructions, hazard warnings.” (acquired in Concept 8)

In this progression, we can find several pitfalls. According to it, a student would be able to retrieve specific information in a document containing cultural specificities (such as a restaurant menu) before being able to read easy notes or recognize basic radicals. This is exacerbated by the fact that an essential skill like 5 is proposed in the very last concept of the path.

Another drawback that emerged from the analysis is that too many skills are clustered within the same concept. Although we devised an algorithm to achieve learning smoothness, a more complex lattice presenting a lower ratio of skills per concept would be extremely useful to fragment the space and give more accurate hints regarding students’ learning combinations.

A consequence of a high ratio of attributes per concept is that the surmise relationships between skills may be too coarse and not mirror reality. Indeed, a statistical analysis showed that, on average, each skill in the domain surmises around 85 skills and is surmised by around 50. These numbers are definitely too large to achieve refined results. Accordingly, testing the assessment process at this stage is pointless, as it is based precisely on the surmise relationships between skills.

The partial lack of logical coherence among the items in the learning sequence and the presence of too coarse surmise relationships will probably be mitigated with the injection of new data into the system, which will complexify the structure of the space and polish the results.

4.3 Russian Data

During the course of the internship, exploring the academic scene involved in our same topic, we came into contact with a research group working at the University of Helsinki that has been developing a language learning platform called *Revita* (see Section 2.5). We started a collaboration involving sharing of information, data, and results. As a consequence, we came into possession of a dataset consisting of response patterns by learners of Russian.

The original dataset comes from *Revita* (Katinskaia, Nouri, and Yangarber, 2018; Hou et al., 2019) and consists of more than 206,000 responses from 205 students to instances of 123 skills spread on five CEFR levels (A1-C1). The full list of skills can be found in Appendix C. In order to be fully exploitable using FCA, we reshaped the data into the format presented in Table 2. This brought us to have a table of dimensions 205×123 with a real number between 0 and 1 in each cell, corresponding to the degree of mastery of skill i by individual j .

We discovered that the dataset is highly unbalanced, with 2,564 missing values. To mitigate this, we filtered out individuals who do not have values for most of the

⁴¹A radical is a graphical component of a Chinese character, often a semantic indicator of the meaning associated with the character.

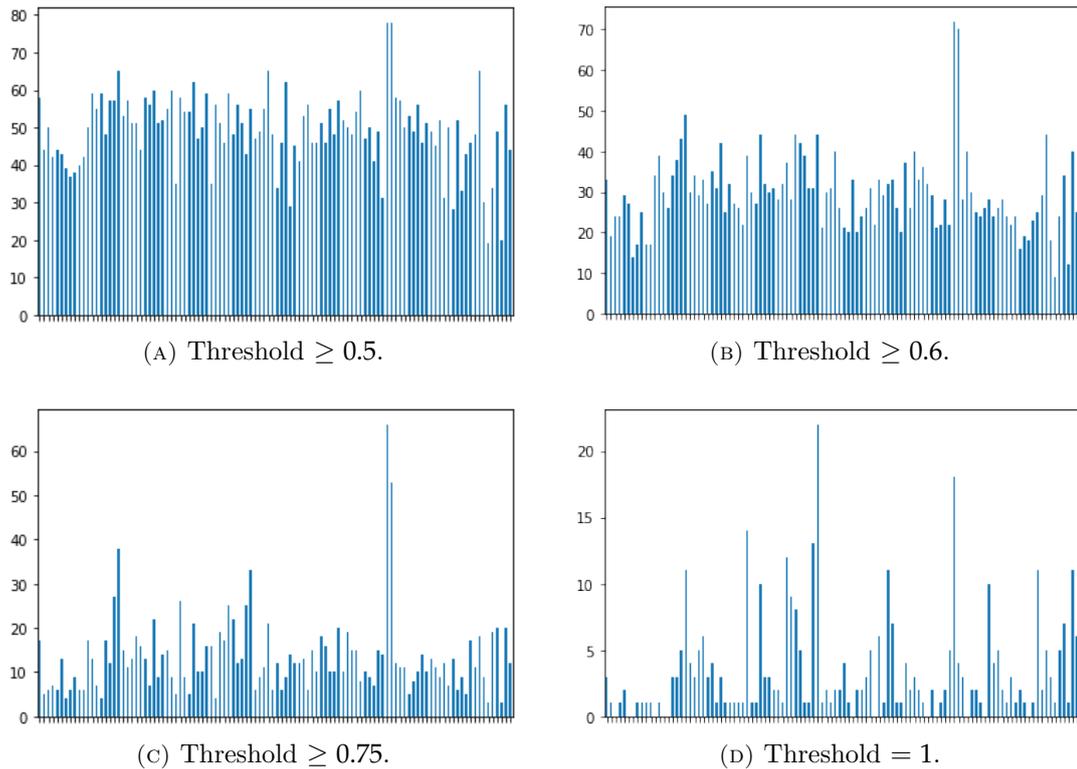


FIGURE 12: Distributions of skills' mastery by students with different acquisition thresholds.

skills. Similarly, we removed the skills for which most of the values were not recorded. At this point, the resulting dataset contains 81 individuals and 108 skills⁴².

The crucial point at this step was to decide the value of the acquisition threshold, for which a skill is considered mastered by a student. Considering that many questions were binary in the test that the individuals undertook, we assumed that testing thresholds lower than 0.5 was not theoretically sound. Thus, we experimented with several thresholds starting from the “random guess” threshold set at 0.5. We eventually picked: 0.5, 0.6, 0.75, and 1.

A closer exploration of the data revealed that, no matter the threshold picked, there is no clear indication that supposedly easier skills are mastered by more students than advanced skills. This behavior can be seen in the graphs in Figure 12. The graphs present on the x -axis the skills ordered by difficulty (easier on the left, more advanced on the right), and on the y -axis the number of students who master those skills. While we expected a monotonically decreasing distribution from left to right following the natural difficulty of the skills, we obtained these jagged plots.

This behavior might be due to several factors, all reflecting inherent challenges related to gathering or tagging the data. Some hypotheses include: (1) the questions posed to the students may not exactly match the difficulty of the skill to which they are related, (2) the problems deriving from the fact that the same skill is sometimes tagged with different CEFR labels, (3) the skills may not be as self-contained as they should be.

⁴²Future work might include the exploration of recommendation systems to predict these missing values, instead of dropping them.

Nonetheless, we decided to set the threshold ≥ 0.6 and input the formal context thus acquired into our learning space. Surprisingly, we discovered that the learning space makes 96 out of 108 skills accessible from the ground level. This result indicates that students with no prior knowledge would theoretically be able to acquire any of these 96 skills without any prerequisites. Clearly, this is not possible in practice. In fact, among these skills, we can find very advanced ones (nine of them are tagged as C1), which should not be accessible at the bottom of the structure. Moreover, as mentioned in Section 4.2.1, an ideal knowledge structure should have few states at the bottom, and here it is not the case.

Next, we computed the best learning path using the same setting as with Mandarin data in Section 4.2.1. The full path can be found in Table 7 in Appendix D. Analyzing the sequence, we can observe that, as opposed to the Mandarin path, a more gradual progression is achieved. Indeed, the ratio of skills per concept is much lower than the one in the Mandarin path. Nevertheless, the learning sequence is not ideal as no evident logical order between easy and advanced skills can be found.

The shortcomings highlighted in this section probably reflect the issues relating to the distribution of skills' mastery pointed out in the graphs in Figure 12. Addressing them might solve or at least mitigate the problems presented in this section, improving the results considerably.

4.3.1 Testing the Assessment Process

Studying one-to-one implications between skills, we found interesting results. Precisely, we found 24 implications, among which 16 seem to be plausible, based on the CEFR tags of the skills. Plausible implications, even if in a modest number, allowed us to test the assessment process.

We took 15 response patterns from the original Revita dataset that were not injected into the learning space, as if we had recruited 15 new students for taking the assessment test (see Section 3.3.2). We know from Revita the actual CEFR level of the 15 students who obtained the response patterns we selected from the dataset. This way, we had a golden label against which we could compare the results of the assessment test.

We got mixed results from this experiment. On a positive note, the number of questions needed to complete the assessment process is, on average, considerably fewer than the number of skills present in the domain. In particular, an average of 31.8 questions ($min = 11$, $max = 50$) was needed to complete the test. Thus, one of the crucial objectives of KST, which is delineating the actual knowledge state of an individual with a limited number of questions, is met.

On a negative note, the knowledge states identified by our tool are mostly the same for all the 15 students. A more thorough analysis of the original data showed that the actual CEFR tags of the individuals are not consistent with the skills mastered by them. Indeed, we found out that students possessing an actual A1 level mastered skills tagged as B2 or even C1. This probably reinforces the hypotheses already put forward in the previous section related to the difficulty of gathering and labeling data of this type. Another reason this happened may be related to the simplicity of the actual assessment process, which is solely based on the calculation of adjacent states and fringes.

4.4 Teachers' Opinions

To lend more weight to our learning space, we interviewed four teachers of Mandarin and Russian and showed them the functionalities and the potential of our tool. This section presents the opinions emerged from the discussions that ensued with them.

The teachers we spoke with were excited about the learning space's potential. All of them would be in favor of testing it in future iterations to experience its evolution. They would also pay for it as a service, provided that it makes their lives easier in planning lessons and finding appropriate native content tailored to the students. In fact, this is an extremely time-consuming task: they mentioned that they need from 1 to 5 hours to outline the program of a single lesson.

As for the learning path, teachers expressed no need for help to define the sequence of skills taught in their lessons. Indeed, this sequence is already predetermined by national programs, and the manuals and institutions they work for already follow these syllabuses. A notable exception was represented by one-on-one tutoring. In this setting, each student has personalized needs that differ from those of their peers. More freedom is allowed, and guidance for both the teacher and the student would be appreciated. In particular, the teacher would need help to find the best learning path towards the final personalized objective, while the student could use the tool to become more autonomous and independent.

Teachers were very interested in implications and prerequisites recognized by the learning space, mainly for the opportunity to identify fine-grained subskills needed to master overarching skills that they are required to teach. However, they believe that its current suggestions were at best incomplete (i.e., other skills were arguably more urgent to learn), and in some cases, downright wrong (i.e., some implied skills bore no relation to the new skill). Nevertheless, they were indulgent when we explained that our input data were limited.

What the teachers valued the most was the potential of the assessment process. Using it, they would have the ability to evaluate learners (frequently and automatically) as they progress. Regular testing would delineate the exact changes in students' knowledge and help identify gaps before they grow. Teachers perceived the process as an opportunity to deal with learners who fall behind, which is regarded as a chronic issue in language classes. The reasons for which students fall behind are diverse (i.e., they lack fundamental skills, or they neglect their homework), yet it is essential to mitigate this issue since students working in groups influence each other, complicating further progress and holding back the rest of their class. One teacher brought up that in classroom settings, the assessment process could help identify groups of students on the same level and assist teachers in differentiating the work.

Teachers also valued the prospect of more accurate evaluation using the learning space, particularly in placement tests, where the assessment is very complex. One teacher drew attention to the administrative and emotional burden for teachers and learners alike of switching classes during language programs.

Finally, teachers were also excited that the learning space could ultimately free language assessment from subjective bias from the evaluators. One teacher was excited about data-driven approaches using learners' corpora to delineate fine-grained skills.

4.5 Discussion

We opened the chapter with three fundamental questions that drove our research experiments. After the analysis of the results obtained in our preliminary experiments, we can try to give some arguments to respond to those questions.

1. *Can KST be applied to language acquisition and teaching?*

Probably only time will tell. At least, the results of the experiments we performed in this section are promising, showing that the learning space is versatile enough to adapt to multiple languages. We indeed created it with Mandarin in mind, but it turned out to work also with Russian data. Moreover, scalable and dynamic methods are vital to prevent the space from being hampered by big formal contexts.

2. *Is a learning space for languages useful to teachers and students?*

Teachers were very excited by the potential of the learning space. In particular, they perceived the learning space as an opportunity to quickly and frequently assess students to identify gaps in their knowledge and improve the quality of one-on-one teaching, which needs to be tailored to the student's needs.

3. *Can we find implications between skills, which would help uncover different learning strategies?*

This seems feasible, but some actions need to be taken for this to be realized successfully. First of all, selecting appropriate skills, no matter the language or the scope, is fundamental. Helpful pointers could be restricting the body of knowledge both horizontally (i.e., focusing on a specific competence, such as reading or writing) and vertically (i.e., focusing on a single CEFR level) and making the skills as much granular as possible. Most importantly, for data-driven approaches, gathering abundant, clean, and correctly tagged data from learners' corpora is crucial to complexify the structure. However, these tasks are known to be complicated and time-consuming.

Chapter 5

Conclusion

In this thesis, we have detailed the conceiving and realization of a learning space for Mandarin acquisition. Leveraging KST and FCA as mathematical foundations, we succeeded in devising a tool that encodes the domain of relevant skills and competences necessary to achieve fluency in Mandarin and orders them into a concept lattice. This mathematical structure helps uncover implications between skills, notably revealing the different learning paths students may follow when acquiring Mandarin. These paths can aid teachers in speeding up testing and in designing lesson plans. Remarkably, our approach is computationally scalable and lightweight, being based on a local exploration of the space.

The remainder of this chapter is organized as follows. In Section 5.1, we will discuss the difficulties we encountered while carrying out the project. After that, in Section 5.2, we will detail future trajectories the project may take and ideas that can be unfolded and developed to show the full potential of this tool. Finally, we will conclude the manuscript with Section 5.3 with some final thoughts about the work presented in this thesis.

5.1 Difficulties

The project presented numerous challenges. In the first place, the construction of the lattice posed difficulties when trying to exploit the tools available online for this use. As we have seen, they cannot handle large datasets, which is essential for our tool. Thus, we were forced to abort the idea of constructing a complete ordered structure. However, this hardship led us to conceive one of the strong points of the tool. Indeed, the local approach based on the exploration of only the parts of the space needed at each search is lightweight and scalable. Using this approach should allow the tool to accommodate vast formal contexts containing thousands of students and skills.

Secondly, the lack of appropriate corpora for our needs forced us to launch a data collection campaign. This campaign entailed adapting existing linguistic assessment criteria to our needs. In particular, crafting skills that would be as self-contained as possible to assess a precise notion and creating instances of those skills that would cover the same notion were demanding tasks. Moreover, correcting skills and instances following several iterations of users' and experts' feedback was time-consuming.

Third, developing the WenTest took twice as long as expected. We faced unforeseeable delays in its development, which postponed the collection of response patterns from test-takers. The deferments were mainly due to making the website more user-friendly and including elements of gamification. Data could have been acquired using less sophisticated means. However, finding people willing to serve as research subjects

proved arduous. Had the users found the interface irritating, we might have had even more trouble recruiting them.

Fourth, working in a team with no prior experience on such a decisively complex subject was challenging. Our topic of study is at the crossroads of several disciplines (i.e., CALL, ITS, language pedagogy, KST, and FCA). Moreover, there is apparently no full-fledged application of KST for language teaching. Starting from these premises, adopting an appropriate standpoint for the research was not effortless. However, we succeeded in doing this by digging deep into the literature of each discipline to find unexplored links between very diverse topics and finally meet the demands of our project. In this respect, it proved helpful to meet the team working on the Revita project, with whom we enjoyed a fruitful collaboration.

Lastly, devising proper evaluations to examine the quality of the tool was extremely difficult. Usually, ITSs, whether intended to aid language teaching and learning or other scholarly topics, are subject to human-based qualitative evaluations, which are notoriously costly and time-consuming. Quantitative evaluations are also possible, but only when sufficient data (i.e., regarding students' learning curves, effectiveness and efficacy, improvement of teaching's easiness) are gathered (Greer and Mark, 2016). Clearly, we do not have access to this type of data yet. However, we managed to devise some preliminary evaluations for single parts of the tool with the few data at our disposal. Moreover, we succeeded in recruiting Mandarin and Russian teachers who could test the tool.

5.2 Future Work

This thesis lays the foundations for using learning spaces for language acquisition. From this starting point, multiple directions can be taken to improve the work that has been proposed in this manuscript. Some ideas for future directions have already been highlighted throughout the thesis. In this final section, we will underline and expand the most significant from a research viewpoint.

First and foremost, the need for a constant flux of new data to improve the system's overall quality is indispensable. This could be accomplished by gathering more response patterns using WenTest or other existing platforms, or finding suitable corpora, i.e., through collaborations with institutions, such as universities or language institutes. When these data are available, more thorough experiments, possibly on powerful machines, could be conducted. This way, more accurate metrics, such as learning curves and support, could be computed to understand if learning spaces can be really employed in language acquisition and discover their actual utility. Another direction that could be taken to mitigate the burden of acquiring a multitude of data for data-driven approaches is exploring the “query to expert” methodology (see Section 2.1) to build the space.

Second, at the time of writing, the structure of our learning space is mostly deterministic. The exploration of more sophisticated algorithms that take into account probabilities is essential for better mitigation of lucky guesses and careless errors, in both data collection and during the assessment process. A first step could be removing or diminishing the number of multiple-choice questions and open up to graded responses. This could also alleviate the process of selection of an acquisition threshold of skills. Then, the feasibility of introducing probabilistic models, such as BLIM (see Section 2.1), BKT (see Section 1.1), or spaced repetition (Dempster, 1989) could

be investigated. More sophistication could also improve our algorithm for achieving learning smoothness, which is now quite naive, being based on the computation of the powerset, and refine the scoring function we use. This way, more adjustability to the learning style of the student may be attained. Finally, the assessment process too would benefit from probabilistic models, given that it is now based essentially only on the computation of adjacent states and fringes.

Lastly, extreme care should be taken in selecting more appropriate and fine-grained skills, using a more psycholinguistic and pedagogical-oriented approach, closely collaborating with experts in the field. This is clearly a complex task, given the non-discrete and ambiguous nature of human languages. Helpful could be splitting the actual learning space into more restricted subspaces, each focused on a specific competence, such as reading, writing, listening, speaking, or a specific HSK or CEFR level. Consequently, more consistent paths and accurate implications would be found.

5.3 Final Words

The work presented in this thesis shows that creating a learning space for language acquisition seems feasible, although considerably arduous. Indeed, we managed to make a prototype version run. However, for a full-fledged realization, much more time and resources are needed. Notably, it is extremely hard to define skills and test questions and acquire and label high-quality data that fulfill the criteria needed to construct a learning space. The internship revealed that the challenge might be unrealistic for a single research team or company, yet collaborations among teams with varied expertise may prove successful.

Nevertheless, our preliminary results suggest that the final objective is attainable, and presenting these results to language teachers revealed that there is demand for this type of tool. Based on our experience, mitigating the inconvenience of acquiring a multitude of data is a priority. Thus, exploring other methods to build the space, such as the query to expert methodology, is vital. Furthermore, splitting the actual learning space into more subspaces is essential so that the body of knowledge would be restricted both competence-wise (listening, reading, writing, or speaking) and level-wise (i.e., A1 to B1 or HSK 1 to HSK 2+).

The foundations for investigating further are laid. In particular, the algorithm is developed and waiting for data, WenTest is online, and teachers have identified their first needs. We leave it to our successors at Naimrod to continue the work that will hopefully culminate in demonstrating the real potential of using learning spaces for language acquisition.

Bibliography

- Ahlgren Reddy, Alison and Marc Harper (2013). “Mathematics placement at the University of Illinois”. In: *Primus* 23.8, pp. 683–702.
- Albert, Dietrich and Josef Lukas (1999). *Knowledge spaces: Theories, empirical research, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Amaral, Luiz A and Detmar Meurers (2011). “On using intelligent computer-assisted language learning in real-life foreign language teaching and learning”. In: *ReCALL* 23.1, pp. 4–24.
- Arasasingham, Ramesh D et al. (2004). “Using knowledge space theory to assess student understanding of stoichiometry”. In: *Journal of Chemical Education* 81.10, p. 1517.
- Baum, Leonard E et al. (1972). “An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes”. In: *Inequalities* 3.1, pp. 1–8.
- Bellassen, Joël (2011). “Is Chinese Eurocompatible? Is the Common Framework common? The Common European Framework of Reference for Languages facing distant languages”. In: *International Symposium Report*, pp. 23–31.
- Birkhoff, Garrett (1940). *Lattice theory*. Vol. 25. American Mathematical Soc.
- Bock, R Darrell and Murray Aitkin (1981). “Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm”. In: *Psychometrika* 46.4, pp. 443–459.
- Cantor, Georg (1874). “Ueber eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen”. In: *Journal für die reine und angewandte Mathematik (in German)*. Vol. 77. Springer, pp. 258–262.
- Chaikalis, Theodore (June 2016). “Graph-Based Software Evolution Analysis of Object-Oriented Systems”. PhD thesis.
- Chen, Yang (2013). “Adaptive Non-graded Assessment Based on Knowledge Space Theory”. In: *2013 IEEE 13th International Conference on Advanced Learning Technologies*. IEEE, pp. 63–64.
- Chiusole, Debora de, Luca Stefanutti, and Andrea Spoto (2017). “A class of k-modes algorithms for extracting knowledge structures from data”. In: *Behavior research methods* 49.4, pp. 1212–1226.
- Corbett, Albert T and John R Anderson (1994). “Knowledge tracing: Modeling the acquisition of procedural knowledge”. In: *User modeling and user-adapted interaction* 4.4, pp. 253–278.
- Cosyn, Eric et al. (2021). “A practical perspective on knowledge space theory: ALEKS and its data”. In: *Journal of Mathematical Psychology* 101, pp. 102–512.
- Davey, Brian A and Hilary A Priestley (2002). *Introduction to lattices and order*. Cambridge university press.
- Dempster, Frank N (1989). “Spacing effects and their implications for theory and practice”. In: *Educational Psychology Review* 1.4, pp. 309–330.

- Doignon, Jean-Paul (1994). “Knowledge spaces and skill assignments”. In: *Contributions to mathematical psychology, psychometrics, and methodology*. Springer, pp. 111–121.
- (2014). “Learning spaces, and how to build them”. In: *International Conference on Formal Concept Analysis*. Springer, pp. 1–14.
- Doignon, Jean-Paul and Jean-Claude Falmagne (1985). “Spaces for the assessment of knowledge”. In: *International journal of man-machine studies* 23.2, pp. 175–196.
- (1999). *Knowledge Spaces*. Berlin, Heidelberg, and New York: Springer-Verlag.
- (2012). *Knowledge spaces*. Berlin: Springer Science & Business Media.
- (2016). “Knowledge spaces and learning spaces”. In: *New handbook of mathematical psychology*, pp. 274–321.
- Dolques, Xavier, Florence Le Ber, and Marianne Huchard (2013). “AOC-posets: a scalable alternative to Concept Lattices for Relational Concept Analysis”. In: *CLA: Concept Lattices and their Applications*, pp. 129–140.
- Dowling, Cornelia E (1993). “Applying the basis of a knowledge space for controlling the questioning of an expert”. In: *Journal of Mathematical Psychology* 37.1, pp. 21–48.
- Elo, Arpad E (1978). *The rating of chessplayers, past and present*. Arco Pub.
- Embretson, Susan E and Steven P Reise (2013). *Item response theory*. Psychology Press.
- Falmagne, Jean-Claude and Jean-Paul Doignon (1988). “A class of stochastic procedures for the assessment of knowledge”. In: *British Journal of Mathematical and Statistical Psychology* 41.1, pp. 1–23.
- (2010). *Learning spaces: Interdisciplinary applied mathematics*. Berlin: Springer Science & Business Media.
- Falmagne, Jean-Claude et al. (1990). “Introduction to knowledge spaces: How to build, test, and search them.” In: *Psychological Review* 97.2, p. 201.
- Falmagne, Jean-Claude et al. (2006). “The assessment of knowledge, in theory and in practice”. In: *Formal concept analysis*. Springer, pp. 61–79.
- Falmagne, Jean-Claude et al. (2013). *Knowledge spaces: Applications in education*. Berlin: Springer Science & Business Media.
- Ganter, Bernhard and Rudolf Wille (2012). *Formal concept analysis: mathematical foundations*. Springer Science & Business Media.
- Greer, Jim and Mary Mark (2016). “Evaluation methods for intelligent tutoring systems revisited”. In: *International Journal of Artificial Intelligence in Education* 26.1, pp. 387–392.
- Hall, Marshall and Richard Bellman (1960). “Combinatorial Analysis”. In: *Proceedings of the Tenth Symposium in Applied Mathematics of the American Mathematical Society*. Ed. by Richard Bellman and Marshall Jr. Hall. Columbia University: American Mathematical Society.
- Hart, Robert (1981). “Language study and the PLATO system.” In: *Studies in language learning* 3.1, pp. 1–24.
- Heift, Trude (2008). “Modeling learner variability in CALL”. In: *Computer Assisted Language Learning* 21.4, pp. 305–321.
- (2010). “Developing an intelligent language tutor”. In: *CALICO journal* 27.3, pp. 443–459.
- Heller, Jürgen et al. (2013). “Recent developments in competence-based knowledge space theory”. In: *Knowledge spaces*. Springer, pp. 243–286.
- Hockemeyer, Cord, Theo Held, and Dietrich Albert (1998). “RATH-A relational adaptive tutoring hypertext WWW-environment based on knowledge space theory”. In: *Computer Aided Learning and Instruction in Science and Engineering. Proceedings*

- of the Fourth International Conference CALISCE'98. Ed. by Christer Alvegard. Goteborg, Sweden: Chalmers University of Technology.
- Hou, Jue et al. (Aug. 2019). "Modeling language learning using specialized Elo rating". In: *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*. Florence, Italy: Association for Computational Linguistics, pp. 494–506. URL: <https://aclanthology.org/W19-4451>.
- Huang, Xudong et al. (2016). "Intelligent tutoring systems work as a math gap reducer in 6th grade after-school program". In: *Learning and Individual Differences* 47, pp. 258–265.
- Kambouri, Maria et al. (1994). "Knowledge assessment: Tapping human expertise by the QUERY routine". In: *International Journal of Human-Computer Studies* 40.1, pp. 119–151.
- Katinskaia, Anisia, Javad Nouri, and Roman Yangarber (May 2018). "Revita: a Language-learning Platform at the Intersection of ITS and CALL". In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA). URL: <https://aclanthology.org/L18-1644>.
- Koppen, Mathieu (1993). "Extracting human expertise for constructing knowledge spaces: An algorithm". In: *Journal of mathematical psychology* 37.1, pp. 1–20.
- Koppen, Mathieu and Jean-Paul Doignon (1990). "How to build a knowledge space by querying an expert". In: *Journal of Mathematical Psychology* 34.3, pp. 311–331.
- Lee, Sanghoon et al. (Sept. 2014). "A Web Application for the Knowledge Space Theory Analysis: KST Web". In: *International Journal of Engineering Research & Technology (IJERT)* 3.
- Levy, Michael (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford University Press.
- Little, David (2007). "The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy". In: *The Modern Language Journal* 91.4, pp. 645–655.
- (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Strasbourg: Council of Europe Publishing.
- Lord, Frederic M (2012). *Applications of item response theory to practical testing problems*. Routledge.
- Lu, Xiaofei and Berlin Chen (2019). *Computational and corpus approaches to Chinese language learning*. Springer.
- Lukas, Josef and Dietrich Albert (1993). "Knowledge assessment based on skill assignment and psychological task analysis". In: *Advances in psychology*. Vol. 101. Elsevier, pp. 139–159.
- Lynch, Collin F et al. (2006). "Defining ill-defined domains; a literature survey". In: *Intelligent Tutoring Systems (ITS 2006): Workshop on Intelligent Tutoring Systems for Ill-Defined Domains*.
- Margaritis, Dimitris (2003). *Learning Bayesian network model structure from data*. Tech. rep. Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.
- Matayoshi, Jeffrey and Eric Cosyn (2018). "Identifying Student Learning Patterns with Semi-Supervised Machine Learning Models". In: *Proceedings of the 26th International Conference on Computers in Education*, pp. 11–20.
- Müller, Cornelia E (1989). "A procedure for facilitating an expert's judgements on a set of rules". In: *Mathematical psychology in progress*. Springer, pp. 157–170.
- Nagata, Noriko (2009). "Robo-Sensei's NLP-based error detection and feedback generation". In: *Calico Journal* 26.3, pp. 562–579.

- Nicotra, Eraldo Francesco and Andrea Spoto (2019). “Connections and Dissimilarities Among Formal Concept Analysis, Knowledge Space Theory and Cognitive Diagnostic Models in a Systemic Perspective”. In: *Systemics of Incompleteness and Quasi-Systems*. Springer, pp. 235–241.
- Nkambou, Roger, Riichiro Mizoguchi, and Jacqueline Bourdeau (2010). *Advances in intelligent tutoring systems*. Vol. 308. Springer Science & Business Media.
- Osswald, Rainer and Wiebke Petersen (2003). “A logical approach to data-driven classification”. In: *Annual Conference on Artificial Intelligence*. Springer, pp. 267–281.
- Pearl, Judea (2009). *Causality*. 2nd ed. Cambridge University Press.
- Petersen, Wiebke (2004). “A set-theoretical approach for the induction of inheritance hierarchies”. In: *Electronic Notes in Theoretical Computer Science* 53, pp. 296–308.
- Poelmans, Jonas et al. (2013). “Formal concept analysis in knowledge processing: A survey on applications”. In: *Expert systems with applications* 40.16, pp. 6538–6560.
- Priss, Uta (2006). “Formal concept analysis in information science”. In: *Annual review of information science and technology* 40.1, pp. 521–543.
- Psotka, Joseph et al. (1988). *Intelligent tutoring systems: Lessons learned*. Psychology Press.
- Rusch, Anja and Rudolf Wille (1996). “Knowledge spaces and formal concept analysis”. In: *Data analysis and information systems*. Ed. by H. H. Bock and W. Polasek. Berlin: Springer, pp. 427–436.
- Samejima, Fumiko (1997). “Graded response model”. In: *Handbook of modern item response theory*. Springer, pp. 85–100.
- Sargin, Anatol and Ali Ünlü (2009). “Inductive item tree analysis: Corrections, improvements, and comparisons”. In: *Mathematical Social Sciences* 58.3, pp. 376–392.
- Schrepp, Martin (1999). “Extracting knowledge structures from observed data”. In: *British Journal of Mathematical and Statistical Psychology* 52.2, pp. 213–224.
- (2003). “A method for the analysis of hierarchical dependencies between items of a questionnaire”. In: *Methods of Psychological Research Online* 19, pp. 43–79.
- Schrepp, Martin and Theo Held (1995). “A simulation study concerning the effect of errors on the establishment of knowledge spaces by querying experts”. In: *Journal of Mathematical Psychology* 39.4, pp. 376–382.
- Singh, Prem Kumar, Cherukuri Aswani Kumar, and Abdullah Gani (2016). “A comprehensive survey on formal concept analysis, its research trends and applications”. In: *International Journal of Applied Mathematics and Computer Science* 26.2.
- Slavuj, Vanja, Božidar Kovačić, and Igor Jugo (2015). “Intelligent tutoring systems for language learning”. In: *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, pp. 814–819.
- Spoto, Andrea, Luca Stefanutti, and Giulio Vidotto (2010). “Knowledge space theory, formal concept analysis, and computerized psychological assessment”. In: *Behavior research methods* 42.1, pp. 342–350.
- (2016). “An iterative procedure for extracting skill maps from data”. In: *Behavior research methods* 48.2, pp. 729–741.
- Stewart, Ian Nicholas (2015). *Galois theory*. CRC press.
- Strok, Fedor and Alexey Neznanov (2010). “Comparing and Analyzing the Computational Complexity of FCA Algorithms”. In: *Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*. SAICSIT '10. Bela Bela, South Africa: Association for Computing Machinery, pp. 417–420. ISBN: 9781605589503. URL: <https://doi.org/10.1145/1899503.1899557>.

- Taagepera, Mare and Ramesh D Arasasingham (2013). “Using knowledge space theory to assess student understanding of chemistry”. In: *Knowledge Spaces*. Springer, pp. 115–128.
- Tafazoli, Dara, Elena Gómez María, and Cristina A Huertas Abril (2019). “Intelligent Language Tutoring System: Integrating Intelligent Computer-Assisted Language Learning Into Language Education”. In: *International Journal of Information and Communication Technology Education (IJICTE)* 15.3, pp. 60–74.
- Villano, Michael (1991). “Computerized knowledge assessment: Building the knowledge structure and calibrating the assessment routine”. PhD thesis. New York University.
- Vu, Anh-Duc (2020). *Modeling knowledge states in language learning*. Helsingin yliopisto.
- Wille, Rudolf (1981). “Restructuring lattice theory: an approach based on hierarchies of concepts”. In: *Ordered Sets. Proceedings of the NATO Advanced Study Institute*. Ed. by Ivan Rival. Springer Netherlands. Banff, Canada, pp. 445–470.
- Yang, Yang et al. (2011). “Automatic dance lesson generation”. In: *IEEE Transactions on Learning Technologies* 5.3, pp. 191–198.
- Zhang, George X and Linda M Li (2019). “Chinese Language Learning and Teaching in the UK”. In: *The Routledge Handbook of Chinese Language Teaching*. Routledge, pp. 565–580.

Appendix A

List of Mandarin Skills

Skill	CEFR	HSK
Can recognise numbers, prices, dates and days of the week, introductions, provided they are delivered in a defined, familiar everyday context.	A0+	HSK0+
Understands the basics of the Chinese phonology.	A0+	HSK0+
Can recognise familiar words/signs accompanied by pictures, such as a picture book using familiar vocabulary.	A0+	HSK1
Can understand very short, simple, instructions used in familiar everyday contexts (e.g. “No parking” , “No food or drink”), especially if there are illustrations.	A0+	HSK1
Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases.	A1	HSK1
Can get an idea of the content of simpler informational material and short, simple descriptions, especially if there is visual support.	A1	HSK1
Can recognise basic sentence patterns to ask a 吗 question.	A1	HSK1
Can understand short, simple dialogue between two persons provided they are delivered slowly and clearly in a defined, familiar everyday context.	A0+	HSK1+
Can follow a short, simple talk about everyday matters provided it is delivered slowly, clearly and with long pauses.	A0+	HSK1+
Can recognise words/signs, names and numbers that they already know in simple, short recordings, provided these are delivered very slowly and clearly.	A0+	HSK1+
Can understand in outline short texts in illustrated stories, provided the images help them to guess at a lot of the content.	A1	HSK1+
Can recognise simple volitive sentences.	A1	HSK1+
Can recognise basic ways to express the ability to do something.	A1+	HSK2
Can recognise times and places in very simple notes and recordings from friends or colleagues (e.g. “Back at 4 o’ clock” or “In the meeting room”).	A1	HSK2
Can find information about places, times and prices on posters, flyers and notices.	A1	HSK2

Can understand words/signs and short sentences in a simple conversation (e.g. between a customer and a salesperson in a shop), provided people communicate very slowly and very clearly.	A1	HSK2
Can understand in outline very simple information being explained in a predictable situation like a guided tour, provided the delivery is very slow and clear and that there are long pauses from time to time.	A1	HSK2
Can understand when someone tells them slowly and clearly where something is, provided the object is in the immediate environment.	A1	HSK2
Can understand store guides (information on which floors departments are on) and directions (e.g. where to find lifts).	A1	HSK2
Can understand phrases and expressions related to areas of most immediate priority (e.g. very basic personal and family information, shopping, local geography, employment).	A1+	HSK2
Can follow in outline short, simple social exchanges, conducted very slowly and clearly.	A2	HSK2
Can understand simple directions on how to get from X to Y, by foot or public transport.	A2	HSK2
Can understand information given in illustrated brochures and maps (e.g. the principal attractions of a city).	A2	HSK2
Can understand simple information about travel and use of public transport (e.g. buses, trains, taxis).	A2	HSK2
Can tell apart pronunciation of quasi-homophonous basic characters.	A1	HSK2
Can recognise basic lexicon expressing time.	A1	HSK2
Can recognise wordings involving means of transportation.	A2	HSK2
Can recognise basic Chinese radicals in characters.	A2	HSK2
Can recognise basic wh-question patterns.	A1	HSK2
Can use demonstrative sentences.	A2	HSK2
Can express the intensity of a verb with adverbs.	A2	HSK2
Can recognise basic expressions of negation.	A1+	HSK2
Can recognise basic question patterns to ask about someone.	A1	HSK2
Can identify ways to shorten expressions and make them more natural.	A1	HSK2
Can identify which interrogative word is most suitable.	A1	HSK2
Understands sentences organizing events on a timeline.	A2	HSK2
Can understand simple everyday signs such as “Parking” , “Station” , “Dining room” , “No smoking” , etc.	A1+	HSK2+
Can recognise concrete information (e.g. places and times) on familiar topics encountered in everyday life.	A1	HSK2+
Understands basic ways to link parts of sentences.	A2	HSK2+
Can recognise basic ways to express obligation.	A2	HSK2+
Can understand short, simple messages sent via social media or e-mail (e.g. proposing what to do, when and where to meet).	A1	HSK2+

Can understand the outline of simple information given in a predictable situation, such as on a guided tour (e.g. “This is where the President lives”).	A2	HSK2+
Can catch the main point in short, clear, simple messages and announcements.	A2	HSK2+
Can understand short, simple personal letters.	A2	HSK2+
Can recognise adverbs expressing the degree of a state.	A2	HSK2+
Can recognise basic structures to express appreciation.	A2	HSK2+
Can recognise basic sentence patterns to emphasize one part of the message.	A2	HSK2+
Can express superlatives.	B1	HSK2+
Understands basic result-focused structures.	A2	HSK2+
Can recognise basic expressions related to alternatives.	A2	HSK2+
Can recognise basic structures expressing causality.	A2	HSK2+
Can recognise basic traditional characters.	A1+	HSK2+
Can recognise basic structures involving auxiliary verbs.	A2	HSK2+
Can recognise everyday, familiar words/signs, provided they are delivered in a clearly defined, familiar everyday context.	A2+	HSK3
Can simply express duration.	A1	HSK3
Can extract important information from short broadcasts (e.g. the weather forecast, concert announcements, sports results), provided people talk clearly.	A2	HSK3
Can find specific, predictable information in simple everyday material such as advertisements, webpages, prospectuses, menus, reference lists, and timetables.	A2	HSK3
Can understand everyday signs and notices, etc. in public places, such as streets, restaurants, railway stations; in workplaces, such as directions, instructions, hazard warnings.	A2	HSK3
Can understand short narratives and descriptions of someone’s life composed in simple language.	A2	HSK3
Can understand short, basic to intermediate messages sent via social media or e-mail (e.g. proposing what to do, when and where to meet).	A2	HSK3
Can understand short conversations conducted slowly and clearly.	A2+	HSK3
Can understand in an interview what people say they do in their free time, what they particularly like doing and what they do not like doing, provided they speak slowly and clearly.	A2+	HSK3
Can understand a simple personal letter, e-mail or post in which the person writing is talking about familiar subjects (such as friends or family) or asking questions on these subjects.	A2+	HSK3
Can find and understand relevant information in everyday material, such as letters, brochures and short official documents.	B1	HSK3
Can recognise expressions related to the location of something.	A2	HSK3

Can recognise adversative transitions between two propositions.	B1	HSK3
Can recognise lexicon concerning moods or states of mind.	A2	HSK3
Can recognise sentences expressing the recurrence of an event.	A2	HSK3
Can recognise kinship terms.	A1	HSK3
Can recognise phonetic components in characters.	A2	HSK3
Can tell apart pronunciation of quasi-homophonous intermediate characters.	B1	HSK3
Can turn an adjective into an adverb.	A2	HSK3
Is able to use measure words related to verbs.	A2	HSK3
Can recognise expressions related to needs.	B1	HSK3
Can recognise intermediate traditional characters.	A2+	HSK3
Can understand basic potential complements.	A2+	HSK3
Can understand basic result complements.	A2+	HSK3
Expressing change of state or completion with 了.	A2+	HSK3
Can recognise wordings expressing politeness.	A2	HSK3
Can recognise intermediate structures to express comparison.	B1	HSK3
Can recognise simple wordings about description.	A2	HSK3
Can understand short, simple texts containing highest frequency vocabulary.	A2	HSK3
Can understand basic instructions on times, dates and numbers, etc., and on routine tasks and assignments to be carried out.	A2	HSK3+
Can understand texts, including web pages, describing people, places, everyday life and culture, etc., provided they use simple language.	A2+	HSK3+
Can recognise measure words.	A2+	HSK3+
Can understand simple, brief instructions, provided they are illustrated and not presented in continuous text.	A2	HSK3+
Can understand straightforward announcements provided the delivery is slow and clear.	A2+	HSK3+
Can understand the most important information contained in short commercials concerning goods and services of interest (e.g. CDs, video games, travel).	A2+	HSK3+
Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or job-related language.	A2+	HSK3+
Can understand the main information in short and simple descriptions of goods in brochures and websites (e.g. portable digital devices, movies, clothes).	A2+	HSK3+
Can understand the main points of short texts dealing with everyday topics (e.g. lifestyle, hobbies, sports, weather).	A2+	HSK3+
Can understand the main points made in short magazine reports or guide entries that deal with concrete everyday topics (e.g. hobbies, sports, leisure activities, animals).	B1	HSK3+

Can understand the main points of news bulletins and simpler recorded material about familiar subjects delivered relatively slowly and clearly.	B1	HSK3+
Can pick out important information about preparation and usage on the labels on foodstuff and medicine.	B1	HSK3+
Has knowledge of the geography where the language is spoken and can make inferences based on this knowledge.	A2+	HSK3+
Can recognise complex expressions related to alternatives.	B1	HSK3+
Can recognise complex structures to express consequence.	B1	HSK3+
Can express situations in progress.	B1+	HSK3+
Can follow detailed directions.	B1	HSK3+
Can recognise structures to express 'towards'.	A2+	HSK3+
Can follow language in a cartoon which use relatively common language but delivered fast.	B1	HSK4
Can understand formal e-mails and letters (e.g. confirmation of a booking or online purchase).	B1	HSK4
Can understand and extract essential information from short, recorded passages dealing with predictable everyday matters that are delivered slowly and clearly.	B1	HSK4
Can generally identify the topic of discussion around them when it is conducted slowly and clearly.	A2+	HSK4
Can understand short narrative text from a video game and infer context from it.	A2+	HSK4
Can understand enough to read short, simple stories and comic strips involving familiar, concrete situations described in high frequency everyday language.	A2+	HSK4
Understands the main vocabulary necessary to navigate on the web.	A2+	HSK4
Can understand the main points of what is said in a straightforward monologue (e.g. a guided tour), provided the delivery is clear and relatively slow.	B1	HSK4
Can recognize and find specific information on a document that includes culture-specific content, like a restaurant menu.	A2+	HSK4
Can recognise complex structures expressing causality.	B1	HSK4
Can recognise complex structures involving concession.	B1	HSK4
Can recognise complex structures using the passive voice.	B2	HSK4
Can recognise complex ways to express obligation.	B1	HSK4
Expressing 'some' with question words.	B1	HSK4
Can recognise complex structures using factitive verbs.	B2	HSK4
Can understand the reduplication of adjectives.	B1	HSK4
Can recognise complex structures to express comparison.	B2	HSK4
Can express 'by means of...?'	B1	HSK4
Can understand simple technical information, such as operating instructions for everyday equipment.	B1	HSK4
Can understand the 是的 structure for emphasizing details.	A2+	HSK4
Can recognise structures related to belonging.	A2+	HSK4
Can understand regulations, for example safety, when expressed in simple language.	B1	HSK4

Can understand what is happening in a photo story (e.g. in a lifestyle magazine) and form an impression of what the characters are like.	A2	HSK4+
Can understand public announcements at airports, stations and on planes, buses and trains, provided these are clearly articulated with minimum interference from [auditory/visual] background noise.	B1	HSK4+
Can understand straightforward personal letters, e-mails or postings giving a relatively detailed account of events and experiences.	B1	HSK4+
Can understand information given in an advertisement picture.	B1	HSK4+
Can understand most short safety instructions, (e.g. on public transport or in manuals for the use of electrical equipment).	B1	HSK4+
Can extract information from shop windows.	B1	HSK4+
Can read maps of transport network and find the most important information.	B1	HSK4+
Can find information on tourist maps.	B1	HSK4+
Can understand simple poems and song lyrics provided these employ straightforward language and style.	B1+	HSK4+
Has knowledge of main events related to the history of China.	B1+	HSK4+
Can understand and navigate websites to fulfil basic tasks, like purchasing a train ticket.	B1	HSK4+
Can collect important information on a website.	B1	HSK4+
Can use structures to express one's opinion.	B2	HSK4+
Can recognise complex words expressing time.	B1+	HSK4+
Can recognise complex structures to express appreciation.	B1	HSK4+
Can recognise complex sentence patterns to emphasize one part of the message.	B1	HSK4+
Can recognise structures using prepositions.	B1+	HSK4+
Can understand large numbers.	B1+	HSK4+
Can recognise structures expressing purpose.	B1	HSK4+
Can recognise complex wordings about description.	B1	HSK4+
Can recognise complex structures to express difference.	B1+	HSK4+
Can understand the main points made in clear standard language or a familiar variety on familiar matters regularly encountered at work, school, leisure, etc., including short narratives.	B1	HSK4+
Is able to read and understand a movie review.	B2	HSK4+
Can recognise the main points of a monologue conducted slowly and clearly.	B1	HSK4+
Can understand descriptions of places, events, explicitly expressed feelings and perspectives in narratives, guides and magazine articles that employ high frequency everyday language.	B1	HSK5
Can understand the main conclusions in a short argued text.	B1	HSK5

Can follow in outline straightforward short talks on familiar topics, provided these are delivered in clearly articulated standard language or a familiar variety.	B1	HSK5
Can understand the main points and important details in stories and other narratives (e.g. a description of a holiday), provided the delivery is slow and clear.	B1	HSK5
Can understand the main points in descriptive notes such as those on museum exhibits and explanatory boards in exhibitions.	B1	HSK5
Can understand a notice about safety issued by an official body.	B2+	HSK5
Can understand a travel diary mainly describing the events of a journey and the experiences and discoveries of the writer.	B2	HSK5
Can follow the plot of stories, simple novels and comics with a clear linear storyline and high frequency everyday language.	B1	HSK5
Can follow much of everyday conversation and discussion, provided it is clearly articulated in standard language or in a familiar variety.	B1+	HSK5
Can identify the main conclusions in clearly signalled argumentative texts.	B1+	HSK5
Can recognise the line of argument in the treatment of the issue presented, though not necessarily in detail.	B1+	HSK5
Can understand instructions and procedures in the form of a continuous text, for instance in a manual, provided they are familiar with the type of process or product concerned.	B1+	HSK5
Can understand children books using narrative language with pictures and pinyin to aid comprehension.	B1+	HSK5
Can make assumption about children books with pinyin.	B1+	HSK5
Can identify the main reasons for and against an argument or idea in a discussion conducted in clear standard language or a familiar variety.	B2	HSK5
Can follow chronological sequence in extended informal discourse, e.g. in a story or anecdote.	B2	HSK5
Can understand announcements and messages on concrete and abstract topics delivered in standard language or a familiar variety at normal speed.	B2	HSK5
Can understand detailed instructions well enough to be able to follow them successfully.	B2	HSK5
Can understand what is said in a personal e-mail or posting even where some colloquial language is used.	B2	HSK5
Can understand articles and reports concerned with contemporary problems in which particular stances or viewpoints are adopted.	B2	HSK5
Can recognise when a text provides factual information and when it seeks to convince readers of something.	B2	HSK5

Can extract specific information from poor quality, [audibly and/or visually] distorted public announcements, e.g. in a station or sports stadium, or on an old recording.	C1	HSK5
Is aware of local distinctions in vocabulary between Taiwan and mainland China.	B1	HSK5
Can express nuances in tone.	B2	HSK5
Can recognise complex traditional characters.	B2	HSK5
Can understand complex complements of movement.	B1+	HSK5
Can recognise complex structures to express contrast.	B1+	HSK5
Understands complex ways to link parts of sentences.	B2	HSK5
Can understand comic strips and set it in the right context.	B2+	HSK5+
Can understand the language of social media.	B2+	HSK5+
Can extract important information from a social media post.	B2+	HSK5+
Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively. Has a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms.	B2	HSK5+
Can follow the essentials of lectures, talks and reports and other forms of academic/professional presentation which are propositionally and linguistically complex.	B2+	HSK5+
Can understand in detail lengthy, complex texts, whether or not these relate to their own area of speciality.	C1	HSK5+
Can read contemporary literary texts and non-fiction produced in the standard form of the language or a familiar variety with little difficulty and with appreciation of implicit meanings and ideas.	C1	HSK5+
Can understand vlog-type content about familiar subjects delivered at a normal speed with cuts.	B2	HSK5+
Can recognise complex structures expressing condition.	B2	HSK5+
Can understand complex negative-affirmative questions.	B2	HSK5+
Can recognize handwritten characters and extract crucial information.	B2+	HSK6
Can read a personal letter written by hand and infer information from the content.	B2+	HSK6
Has knowledge of classical Chinese.	B2+	HSK6
Can keep up with an animated conversation between proficient users of the target language.	C1	HSK6
Can understand the point of view expressed on topics that are of current interest or that relate to their specialised field, provided the talk is delivered in standard language or a familiar variety.	B2+	HSK6
Can understand recordings in the standard form of the language likely to be encountered in social, professional or academic life and identify viewpoints and attitudes as well as the information content.	B2+	HSK6
Can recognise a wide range of idiomatic expressions and colloquialisms, appreciating register shifts.	C1	HSK6

Can understand a wide range of recorded and broadcast material, including some non-standard usage, and identify finer points of detail including implicit attitudes and relationships between people.	C1	HSK6
Can understand slang, idiomatic expressions and jokes.	C1	HSK6
Can understand content dealing with cultural specificities.	C1	HSK6
Can understand articles dealing with complex issues including politics, international affairs, and economy.	C1	HSK6

TABLE 4: Full list of Mandarin skills with corresponding HSK and CEFR tags, ordered by increasing HSK level.

Appendix B

Best Mandarin Path

Concept 1

Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases.

Can recognise words/signs, names and numbers that they already know in simple, short recordings, provided these are delivered very slowly and clearly.

Concept 2

Can recognise familiar words/signs accompanied by pictures, such as a picture book using familiar vocabulary.

Concept 3

Can recognise basic ways to express the ability to do something.

Concept 4

Can recognise numbers, prices, dates and days of the week, introductions, provided they are delivered in a defined, familiar everyday context.

Can get an idea of the content of simpler informational material and short, simple descriptions, especially if there is visual support.

Can find information about places, times and prices on posters, flyers and notices.

Can recognise concrete information (e.g. places and times) on familiar topics encountered in everyday life.

Can understand short, simple dialogue between two persons provided they are delivered slowly and clearly in a defined, familiar everyday context.

Can recognise basic expressions of negation.

Can express the intensity of a verb with adverbs.

Can understand straightforward announcements provided the delivery is slow and clear.

Can recognise basic sentence patterns to ask a 吗 question.

Can follow a short, simple talk about everyday matters provided it is delivered slowly, clearly and with long pauses.

Can tell apart pronunciation of quasi-homophonous basic characters.

Can recognise basic question patterns to ask about someone.

Can understand information given in illustrated brochures and maps (e.g. the principal attractions of a city).

Can follow in outline short, simple social exchanges, conducted very slowly and clearly.

Can understand in outline very simple information being explained in a predictable situation like a guided tour, provided the delivery is very slow and clear and that there are long pauses from time to time.

Can recognise basic traditional characters.

Can find and understand relevant information in everyday material, such as letters, brochures and short official documents.

Can simply express duration.

- Can pick out important information about preparation and usage on the labels on foodstuff and medicine.
- Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or job-related language.
- Can recognize and find specific information on a document that includes culture-specific content, like a restaurant menu.
- Can understand a wide range of recorded and broadcast material, including some non-standard usage, and identify finer points of detail including implicit attitudes and relationships between people.
- Can recognise basic lexicon expressing time.
- Can understand very short, simple, instructions used in familiar everyday contexts (e.g. “No parking”, “No food or drink”), especially if there are illustrations.
- Can understand in outline short texts in illustrated stories, provided the images help them to guess at a lot of the content.
- Can understand simple everyday signs such as “Parking”, “Station”, “Dining room”, “No smoking”, etc.
- Can recognise everyday, familiar words/signs, provided they are delivered in a clearly defined, familiar everyday context.
- Can understand basic potential complements.
- Expressing change of state or completion with 了.
- Can recognise lexicon concerning moods or states of mind.
- Can understand basic result complements.
- Can recognise phonetic components in characters.
- Can identify which interrogative word is most suitable.
- Understands sentences organizing events on a timeline.
- Can find specific, predictable information in simple everyday material such as advertisements, webpages, prospectuses, menus, reference lists, and timetables.
- Can understand short narratives and descriptions of someone’s life composed in simple language.
- Can understand the main points of news bulletins and simpler recorded material about familiar subjects delivered relatively slowly and clearly.
- Can understand the main points made in short magazine reports or guide entries that deal with concrete everyday topics (e.g. hobbies, sports, leisure activities, animals).
- Can understand the main points of short texts dealing with everyday topics (e.g. lifestyle, hobbies, sports, weather).
- Can understand basic instructions on times, dates and numbers, etc., and on routine tasks and assignments to be carried out.
- Can recognise times and places in very simple notes and recordings from friends or colleagues (e.g. “Back at 4 o’clock” or “In the meeting room”).
- Can understand the point of view expressed on topics that are of current interest or that relate to their specialised field, provided the talk is delivered in standard language or a familiar variety.
- Can understand content dealing with cultural specificities.
- Can read maps of transport network and find the most important information.
- Can understand simple poems and song lyrics provided these employ straightforward language and style.
- Can generally identify the topic of discussion around them when it is conducted slowly and clearly.
- Can understand what is happening in a photo story (e.g. in a lifestyle magazine) and form an impression of what the characters are like.

Can recognise adverbs expressing the degree of a state.
 Can recognise basic structures to express appreciation.
 Can understand phrases and expressions related to areas of most immediate priority (e.g. very basic personal and family information, shopping, local geography, employment).
 Can understand store guides (information on which floors departments are on) and directions (e.g. where to find lifts).
 Can understand words/signs and short sentences in a simple conversation (e.g. between a customer and a salesperson in a shop), provided people communicate very slowly and very clearly.

Concept 5

Can recognise simple volitive sentences.
 Understands basic result-focused structures.
 Can tell apart pronunciation of quasi-homophonous intermediate characters.
 Can recognise intermediate traditional characters.
 Can understand the main points and important details in stories and other narratives (e.g. a description of a holiday), provided the delivery is slow and clear.
 Can recognise basic wh-question patterns.
 Can use demonstrative sentences.
 Can identify ways to shorten expressions and make them more natural.
 Can recognise basic sentence patterns to emphasize one part of the message.
 Understands basic ways to link parts of sentences.
 Can understand short, simple messages sent via social media or e-mail (e.g. proposing what to do, when and where to meet).
 Can catch the main point in short, clear, simple messages and announcements.
 Can recognise basic ways to express obligation.
 Can recognise intermediate structures to express comparison.
 Can understand in an interview what people say they do in their free time, what they particularly like doing and what they do not like doing, provided they speak slowly and clearly.
 Can extract important information from short broadcasts (e.g. the weather forecast, concert announcements, sports results), provided people talk clearly.
 Can understand short, simple texts containing highest frequency vocabulary.
 Can recognise complex structures to express consequence.
 Can understand the most important information contained in short commercials concerning goods and services of interest (e.g. CDs, video games, travel).
 Can understand simple, brief instructions, provided they are illustrated and not presented in continuous text.
 Can understand the 是的 structure for emphasizing details.
 Understands the main vocabulary necessary to navigate on the web.
 Can understand short narrative text from a video game and infer context from it.
 Can recognise structures expressing purpose.
 Can understand the main points made in clear standard language or a familiar variety on familiar matters regularly encountered at work, school, leisure, etc., including short narratives.
 Can recognise complex structures to express difference.
 Can understand the main points of what is said in a straightforward monologue (e.g. a guided tour), provided the delivery is clear and relatively slow.
 Can understand straightforward personal letters, e-mails or postings giving a relatively detailed account of events and experiences.

Can extract specific information from poor quality, [audibly and/or visually] distorted public announcements, e.g. in a station or sports stadium, or on an old recording.
Can understand and navigate websites to fulfil basic tasks, like purchasing a train ticket.

Can understand articles and reports concerned with contemporary problems in which particular stances or viewpoints are adopted.

Can collect important information on a website.

Can understand the main conclusions in a short argued text.

Can follow chronological sequence in extended informal discourse, e.g. in a story or anecdote.

Can identify the main reasons for and against an argument or idea in a discussion conducted in clear standard language or a familiar variety.

Can follow much of everyday conversation and discussion, provided it is clearly articulated in standard language or in a familiar variety.

Can understand complex negative-affirmative questions.

Can understand vlog-type content about familiar subjects delivered at a normal speed with cuts.

Concept 6

Can recognise wordings involving means of transportation.

Can recognise complex ways to express obligation.

Can understand short conversations conducted slowly and clearly.

Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively.

Has a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms.

Can understand instructions and procedures in the form of a continuous text, for instance in a manual, provided they are familiar with the type of process or product concerned.

Can understand the main points in descriptive notes such as those on museum exhibits and explanatory boards in exhibitions.

Can understand public announcements at airports, stations and on planes, buses and trains, provided these are clearly articulated with minimum interference from [auditory/visual] background noise.

Can understand information given in an advertisement picture.

Can recognise complex structures involving concession.

Can understand when someone tells them slowly and clearly where something is, provided the object is in the immediate environment.

Concept 7

Can recognise basic Chinese radicals in characters.

Can understand recordings in the standard form of the language likely to be encountered in social, professional or academic life and identify viewpoints and attitudes as well as the information content.

Can understand children books using narrative language with pictures and pinyin to aid comprehension.

Can extract information from shop windows.

Is able to read and understand a movie review.

Can understand and extract essential information from short, recorded passages dealing with predictable everyday matters that are delivered slowly and clearly.

Can recognise expressions related to the location of something.

Can understand short, basic to intermediate messages sent via social media or e-mail (e.g. proposing what to do, when and where to meet).

Can recognise wordings expressing politeness.

Can recognise simple wordings about description.

Can understand short, simple personal letters.

Can express superlatives.

Understands the basics of the Chinese phonology.

Can recognise a wide range of idiomatic expressions and colloquialisms, appreciating register shifts.

Can extract important information from a social media post.

Can understand articles dealing with complex issues including politics, international affairs, and economy.

Can recognise complex structures expressing condition.

Can follow the essentials of lectures, talks and reports and other forms of academic/professional presentation which are propositionally and linguistically complex.

Can understand announcements and messages on concrete and abstract topics delivered in standard language or a familiar variety at normal speed.

Can understand a travel diary mainly describing the events of a journey and the experiences and discoveries of the writer.

Can understand most short safety instructions, (e.g. on public transport or in manuals for the use of electrical equipment).

Can recognise basic structures expressing causality.

Can recognise structures to express ‘towards’.

Can understand texts, including web pages, describing people, places, everyday life and culture, etc., provided they use simple language.

Can understand simple technical information, such as operating instructions for everyday equipment.

Can turn an adjective into an adverb.

Can recognise kinship terms.

Can follow the plot of stories, simple novels and comics with a clear linear storyline and high frequency everyday language.

Can understand enough to read short, simple stories and comic strips involving familiar, concrete situations described in high frequency everyday language.

Can use structures to express one’s opinion.

Can understand what is said in a personal e-mail or posting even where some colloquial language is used.

Can recognise complex words expressing time.

Can understand the main information in short and simple descriptions of goods in brochures and websites (e.g. portable digital devices, movies, clothes).

Concept 8

Can recognise basic structures involving auxiliary verbs.

Can follow in outline straightforward short talks on familiar topics, provided these are delivered in clearly articulated standard language or a familiar variety.

Has knowledge of main events related to the history of China.

Can understand large numbers.

Can recognise complex structures using factitive verbs.

Can recognise measure words.

Can understand everyday signs and notices, etc. in public places, such as streets, restaurants, railway stations; in workplaces, such as directions, instructions, hazard warnings.

- Can understand a simple personal letter, e-mail or post in which the person writing is talking about familiar subjects (such as friends or family) or asking questions on these subjects.
- Can recognise adversative transitions between two propositions.
- Can recognise basic expressions related to alternatives.
- Can understand detailed instructions well enough to be able to follow them successfully.
- Can understand formal e-mails and letters (e.g. confirmation of a booking or online purchase).
- Can understand a notice about safety issued by an official body.
- Can understand complex complements of movement.
- Can recognise structures using prepositions.
- Has knowledge of the geography where the language is spoken and can make inferences based on this knowledge.
- Can express situations in progress.
- Can recognise complex expressions related to alternatives.
- Can recognise sentences expressing the recurrence of an event.
- Can recognise structures related to belonging.
- Can find information on tourist maps.
- Can express ‘by means of...’.
- Can understand simple information about travel and use of public transport (e.g. buses, trains, taxis).
- Can recognise complex structures to express contrast.
- Expressing ‘some’ with question words.
- Can recognise expressions related to needs.
- Can understand simple directions on how to get from X to Y, by foot or public transport.
- Can understand the outline of simple information given in a predictable situation, such as on a guided tour (e.g. “This is where the President lives”).
- Can recognise the main points of a monologue conducted slowly and clearly.
- Can recognise when a text provides factual information and when it seeks to convince readers of something.
- Is aware of local distinctions in vocabulary between Taiwan and mainland China.
- Can make assumption about children books with pinyin.
- Can read contemporary literary texts and non-fiction produced in the standard form of the language or a familiar variety with little difficulty and with appreciation of implicit meanings and ideas.
- Can recognise the line of argument in the treatment of the issue presented, though not necessarily in detail.
- Can understand the language of social media.
- Can keep up with an animated conversation between proficient users of the target language.
- Can recognise complex structures to express comparison.
- Can identify the main conclusions in clearly signalled argumentative texts.
- Can recognise complex structures using the passive voice.
- Can understand the reduplication of adjectives.
- Can express nuances in tone.
- Can understand regulations, for example safety, when expressed in simple language.
- Can understand in detail lengthy, complex texts, whether or not these relate to their own area of speciality.

Can follow detailed directions.
Can recognise complex structures to express appreciation.
Can recognise complex sentence patterns to emphasize one part of the message.
Is able to use measure words related to verbs.
Can recognise complex structures expressing causality.
Can recognise complex wordings about description.
Has knowledge of classical Chinese.
Can read a personal letter written by hand and infer information from the content.
Understands complex ways to link parts of sentences.
Can understand descriptions of places, events, explicitly expressed feelings and perspectives in narratives, guides and magazine articles that employ high frequency everyday language.
Can recognise complex traditional characters.
Can understand slang, idiomatic expressions and jokes.
Can recognize handwritten characters and extract crucial information.
Can follow language in a cartoon which use relatively common language but delivered fast.
Can understand comic strips and set it in the right context.

TABLE 5: Best path identified in the knowledge structure in Figure 11 built from the data gathered from WenTest.

Appendix C

List of Russian Skills

Skill	CEFR
Adjective government.	A1
Animate noun object.	A1
Aspect, expression of action completed in the past.	A1
Declension of adjectives. Type 'хорошего'.	A1
Fleeting vowel in genitive plural.	A1
Genitive plural.	A1
Genitive plural of Pluralia tantum words.	A1
I conjugation. Type 'плакать'.	A1
I conjugation. Type 'рисовать'.	A1
I declension. Fleeting vowels and alternations 'я, е, ё / й (заяц-зайца, заём-займа)'. I declension. Type 'адрес-адреса'.	A1
I declension. Type 'дерево-деревья'.	A1
I declension. Type 'карандаш'.	A1
I declension. Type 'музей-музеи, воробей-воробьи'.	A1
I declension. Type 'солдат-много солдат, сапог-пара сапог'.	A1
II conjugation. Type 'молчать'.	A1
II declension. Type 'армия'.	A1
II declension. Type 'статья'.	A1
II declension. Type 'на -ня'.	A1
Imperative.	A1
Nouns ending in '-у/-ю' in prepositional singular and '-а' in nominative plural.	A1
Nouns with prepositions 'в/на' ending in '-у/-ю' in prepositional singular.	A1
Nouns. Type 'нож-ножом, сторож-сторожем'.	A1
Nouns. Type A: stress on the stem.	A1
Nouns. Type B, B1, B2: stress on the ending.	A1
Nouns. Type C, C1: stress on the stem in singular, on the ending in plural.	A1
Nouns. Type D, D1: stress on the ending in singular, on the stem in plural.	A1
Phrase 'уметь, нравиться, любить,...' + infinitive.	A1
Predicative adverbs and their government. Necessity, possibility, impossibility.	A1
Regular verbs with vowel alternation.	A1
Sentences with dative subject: 'Кате 25 лет'.	A1

Verb. Case and preposition government.	A1
Word order in sentences introducing direct quotations.	A1
Adjective government.	A2
Aspect, expression of action completed in the past.	A2
Constructions with cardinal numerals.	A2
Dative subject + impersonal-reflexive verb: 'мне нравится / кажется / пришлось'.	A2
Factual meaning of verbs.	A2
Frequent prefixed verbs of motion + prepositional constructions.	A2
I declension. Type '-анин/-янин, -ин'.	A2
I declension. Type 'санаторий'.	A2
Inception of action.	A2
Negative sentences.	A2
Noun. Case and preposition government.	A2
Nouns. Type C, C1: stress on the stem in singular, on the ending in plural.	A2
Phrase 'уметь, нравиться, любить,...' + infinitive.	A2
Predicative adverbs and their government. Existence, state, time.	A2
Predicative adverbs and their government. Necessity, possibility, impossibility.	A2
Proximal future.	A2
Unprefixed verbs of motion.	A2
Verb. Case and preposition government.	A2
Word order in sentences introducing direct quotations.	A2
Adjective government.	B1
Aspect, expression of action completed in the past.	B1
Aspect, expression of capability/incapability. ('Тебе этого не понимать/понять!').	B1
Cardinal numbers. 'Сто' vs. 'пятьсот, шестьсот, семьсот, девятьсот'.	B1
Collocations.	B1
Constructions with cardinal numerals.	B1
Constructions with collective numerals.	B1
Expressions of time, place and manner. Preposition, free expressions.	B1
Factual meaning of verbs.	B1
Iterative / potential iterative / qualities.	B1
Lexicology. Lexical semantics.	B1
Nouns. Type B, B1, B2: stress on the ending.	B1
Nouns. Type D, D1: stress on the ending in singular, on the stem in plural.	B1
Ordinal numbers. Type 'третий'.	B1
Passive and its relation to indefinite-personal sentences (equivalent of Finnish impersonal passive).	B1
Phrase 'пора, скорее' + infinitive.	B1
Place of gerunds in the sentence.	B1
Place of negation in the sentence.	B1
Place of participles in the sentence.	B1
Predicative adverbs and their government. Existence, state, time.	B1
Quantifiers. Collective quantifiers in oblique cases.	B1

Resultative.	B1
Verb. Case and preposition government.	B1
Accusative/ergative subject + Impersonal verb: 'Васю тошнит'.	B2
Adjective government.	B2
Аспект, expression of capability/incapability. ('Тебе этого не понимать/понять!').	B2
Cardinal numbers. 'Сто' vs. 'пятьсот, шестьсот, семьсот, девятьсот'.	B2
Constructions with cardinal numerals.	B2
Dative subject + adverb: 'мне (стало) плохо/нужно/скучно'.	B2
Dative subject + impersonal verb: 'мне идет/надоело/везет/хватит'.	B2
Dative subject + impersonal-reflexive verb: 'мне нравится / кажется / пришлось'.	B2
Expression of duration - 'за какое время'.	B2
Genesis-personal sentences.	B2
Impersonal sentences. Infinitival sentences. Sentences, with negative pronouns and adverbs.	B2
Impersonal sentences. Infinitival sentences. Subordinate sentences ('если / прежде чем' + infinitive).	B2
Iterative / potential iterative / qualities.	B2
Lexicology. Lexical semantics.	B2
Negation 'не' + infinitive.	B2
Negative constructions with predicative 'be' (and synonyms), genitive of negation.	B2
Noun. Case and preposition government.	B2
Nouns. Type B, B1, B2: stress on the ending.	B2
Nouns. Type C, C1: stress on the stem in singular, on the ending in plural.	B2
Phrase 'забыть, успеть, удалиться' + infinitive.	B2
Phrase 'уметь, нравиться, любить,...' + infinitive.	B2
Place of adverbs of time, place and manner.	B2
Place of pronouns in a phrase.	B2
Predicative adverbs and their government. Necessity, possibility, impossibility.	B2
Prep + genitive subject + Impersonal verb: 'У меня болит голова/шумит в голове'.	B2
Preterite. Type 'исчезнуть'.	B2
Quantifiers. Collective quantifiers 'оба, обе'.	B2
Second person singular imperative in conditionals ('если').	B2
Sentences of type 'Знаю его как врача'.	B2
Stress: forms of verbal preterite.	B2
Stress: participles.	B2
Stress: short adjectives.	B2
Usage of 'сам' and 'один'.	B2
Verb. Case and preposition government.	B2
Collocations.	C1
Dative subject + impersonal-reflexive verb: 'мне нравится / кажется / пришлось'.	C1
Expressions of time, place and manner. Preposition, free expressions.	C1
Lexicology. Coordination of words.	C1

Lexicology. Lexical semantics.	C1
Noun. Case and preposition government.	C1
Place of negation in the sentence.	C1
Possessive adjectives. Type 'лисий'.	C1
Possessive adjectives. Type 'мамин'.	C1
Predicative adverbs and their government. Necessity, possibility, impossibility.	C1
Sentences of type 'Быть грозе'.	C1
Sentences of type 'Лодку унесло ветром'.	C1
Verb. Case and preposition government.	C1
Word order in sentences introducing direct quotations.	C1

TABLE 6: Full list of Russian skills with corresponding CEFR tags, ordered by increasing CEFR level.

Appendix D

Best Russian Path

	<i>Concept 0</i>
B2	Place of adverbs of time, place and manner.
	<i>Concept 1</i>
B2	Place of pronouns in a phrase.
	<i>Concept 2</i>
A1	Sentences with dative subject: 'Кате 25 лет'.
	<i>Concept 3</i>
C1	Sentences of type 'Лодку унесло ветром'.
	<i>Concept 4</i>
A2	Phrase 'уметь, нравиться, любить,...' + infinitive.
	<i>Concept 5</i>
B1	Place of negation in the sentence.
	<i>Concept 6</i>
B1	Predicative adverbs and their government. Existence, state, time.
	<i>Concept 7</i>
A1	I conjugation. Type 'рисовать'.
	<i>Concept 8</i>
B2	Dative subject + adverb: 'мне (стало) плохо/нужно/скучно'.
	<i>Concept 9</i>
B1	Expressions of time, place and manner. Preposition, free expressions.
A1	Animate noun object.
	<i>Concept 10</i>
B1	Verb. Case and preposition government.
B2	Expression of duration - 'за какое время'.
	<i>Concept 11</i>
A1	Nouns with prepositions 'в/на' ending in '-у/-ю' in prepositional singular.
A2	Predicative adverbs and their government. Existence, state, time.
A1	Predicative adverbs and their government. Necessity, possibility, impossibility.
B2	Dative subject + impersonal-reflexive verb: 'мне нравится / кажется / пришлось'.
B2	Dative subject + impersonal verb: 'мне идет/надоело/везет/хватит'.
	<i>Concept 12</i>
B1	Passive and its relation to indefinite-personal sentences (equivalent of Finnish impersonal passive).
A2	Verb. Case and preposition government.
	<i>Concept 13</i>
A1	Nouns. Type A: stress on the stem.
B1	Constructions with cardinal numerals.

B2_Impersonal sentences. Infinitival sentences. Sentences, with negative pronouns and adverbs.

C1_Expressions of time, place and manner. Preposition, free expressions.

Concept 14

A1_Nouns. Type B, B1, B2: stress on the ending.

A1_Fleeting vowel in genitive plural.

B2_Impersonal sentences. Infinitival sentences. Subordinate sentences ('если / прежде чем' + infinitive).

A1_Declension of adjectives. Type 'хорошего'.

B1_Adjective government.

B1_Phrase 'пора, скорее' + infinitive.

B2_Negation 'не' + infinitive.

B2_Sentences of type 'Знаю его как врача'.

Concept 15

B1_Quantifiers. Collective quantifiers in oblique cases.

B1_Iterative / potential iterative / qualities.

A1_Adjective government.

B2_Lexicology. Lexical semantics.

B1_Nouns. Type B, B1, B2: stress on the ending.

Concept 16

B1_Factual meaning of verbs.

C1_Noun. Case and preposition government.

Concept 17

B1_Ordinal numbers. Type 'третий'.

A1_Genitive plural.

C1_Word order in sentences introducing direct quotations.

A2_Unprefixed verbs of motion.

A1_Nouns. Type D, D1: stress on the ending in singular, on the stem in plural.

A1_Regular verbs with vowel alternation.

A1_I conjugation. Type 'плакать'.

A2_Proximal future.

A1_I declension. Type 'карандаш'.

C1_Sentences of type 'Быть грозе'.

C1_Place of negation in the sentence.

A1_II conjugation. Type 'молчать'.

A2_Negative sentences.

B1_Cardinal numbers. 'Сто' vs. 'пятьсот, шестьсот, семьсот, девятьсот'.

A1_II declension. Type 'армия'.

B1_Constructions with collective numerals.

A1_I declension. Fleeting vowels and alternations 'я, е, ё / й (заяц-зайца, заём-займа)'.
B1_Collolocations.

B2_Verb. Case and preposition government.

B2_Stress: forms of verbal preterite.

B2_Stress: participles.

B2_Usage of 'сам' and 'один'.

C1_Lexicology. Lexical semantics.

B2_Quantifiers. Collective quantifiers 'оба, обе'.

B2_Negative constructions with predicative 'be' (and synonyms), genitive of negation.

B2_Phrase ‘уметь, нравиться, любить,...’ + infinitive.
 B2_Generis-personal sentences.
 A1_I declension. Type ‘дерево-деревья’.
 C1_Collolocations.
 B2_Stress: short adjectives.
 A2_Factual meaning of verbs.
 B1_Nouns. Type D, D1: stress on the ending in singular, on the stem in plural.
 A1_II declension. Type ‘статья’.
 B2_Aspect, expression of capability/incapability. (‘Тебе этого не понимать /
 понять!’).
 C1_Possessive adjectives. Type ‘лисий’.
 B2_Second person singular imperative in conditionals (‘если’).
 A2_I declension. Type ‘-анин/-янин, -ин’.
 B2_Accusative/ergative subject + Impersonal verb: ‘Васю тошнит’.
 A1_Word order in sentences introducing direct quotations.
 B2_Phrase ‘забыть, успеть, удался’ + infinitive.

Concept 18

A2_Inception of action.
 B2_Prep + genitive subject + Impersonal verb: ‘У меня болит голова/шумит в
 голове’.
 B1_Lexicology. Lexical semantics.
 B1_Resultative.
 A2_Frequent prefixed verbs of motion + prepositional constructions.
 A1_I declension. Type ‘адрес-адреса’.
 B2_Noun. Case and preposition government.
 A2_I declension. Type ‘санаторий’.
 B2_Adjective government.
 B2_Iterative / potential iterative / qualities.
 C1_Verb. Case and preposition government.
 A1_Genitive plural of Pluralia tantum words.
 C1_Possessive adjectives. Type ‘мамин’.
 A2_Nouns. Type C, C1: stress on the stem in singular, on the ending in plural.
 A1_Imperative.
 A1_I declension. Type ‘музей-музеи, воробей-воробьи’.
 C1_Lexicology. Coordination of words.
 A1_I declension. Type ‘солдат-много солдат, сапог-пара сапог’.
 A1_Nouns. Type ‘нож-ножом, сторож-сторожем’.
 A1_II declension. Type ‘на -ня’.
 B1_Place of participles in the sentence.
 A1_Nouns ending in ‘-у/-ю’ in prepositional singular and ‘-а’ in nominative plural.
 A2_Word order in sentences introducing direct quotations.
 A1_Nouns. Type C, C1: stress on the stem in singular, on the ending in plural.
 B1_Aspect, expression of capability/incapability. (‘Тебе этого не понимать /
 понять!’).
 B1_Place of gerunds in the sentence.
 B1_Aspect, expression of action completed in the past.

Concept 19

B2_Preterite. Type ‘исчезнуть’.

Concept 20

B2_Cardinal numbers. 'Сто' vs. 'пятьсот, шестьсот, семьсот, девятьсот'.

TABLE 7: Best path identified from Revita's data.