

Digital Models of Metrics for Distance Assessment of Foreign Language Proficiency

Kaftannikov I. L.¹ Tsytovich M. V.^{2,*}

¹School of Electronic Engineering and Computer Science, South Ural State University, Chelyabinsk, 454000, Russia

²Institute of Linguistics and International Communications, South Ural State University, Chelyabinsk, 454000, Russia

* Corresponding author. Email: Maria.tsytovich@gmail.com

ABSTRACT

The rapid development of global civilization processes determines new requirements for the education. At the same time, the widespread introduction of distance learning during the Covid 19 pandemic revealed many problems: problems with direct distance interaction between students and teachers and organizational ones. On the one hand, new broad ranges of requirements for the knowledge and skills of employees are generated, especially in the distance communication sphere, on the other hand, the globalization of media systems leads to a global transformation of educational structures. Therefore, there is a growing awareness of the need in the shift from full-time expert systems for assessing knowledge and skills to automatic digital testing systems. The paper presents an approach based on the creation of digital representations of the full-time expert assessment components, which allows obtaining structural digital assessment schemes with a minimum use of speech recognition tools.

The limitation of the study is the "English language acquisition" subject area with the analysis of knowledge and skills testing at A2 (pre-intermediate) level as the most common one. Methods have been developed for creating digitalization metrics for speaking skill assessment. The developed structural means of transformation of traditional assessment metrics into digital ones are proposed and can be used for checking the main communication component: grammatical and lexical resources, mastering syntax and semantics in a dialogue mode and describing everyday situations, and using phrasal verbs. The substantiation of the possibility to use the proposed approach for automatic distance digital assessment of the foreign language proficiency level is made.

Keywords: *aassessment, metrics, digitalization, foreign language proficiency, Covid - 19, digital models*

1. INTRODUCTION

Currently the world community is witnessing the rapid development of global civilization processes in economy, technology, communication, and etc.

The dynamics of these processes, in turn, determines the transformation and dynamism of the educational environment. Nowadays, on the one hand, new extensive ranges of requirements for the knowledge and skills of workers are generated, especially in the distance communication sphere, on the other hand, the globalization of media systems leads to a global transformation of educational structures and the formation of new teaching methods.

The transformation processes have manifested themselves very clearly and continue to have a significant impact on society with the advent of the corona virus pandemic - Covid-19. Local epidemic factors caused by different types of viruses appeared much earlier (avian and swine flu, atypical pneumonia, Ebola fever, etc.). However, the scale and speed of the Covid-19 spread, assessed as a pandemic, has led to more than just massive medical efforts. There is a need for the widespread use of

teleworking organizations, monitoring and evaluating the results and productivity of work. Similar problems, even in more concentrated form, have to be solved in the field of education.

Moreover, the epidemiologists, economic analysts, and educators at various levels claim that the processes of remote interaction will intensify gradually [1-2].

Covid - 19 has shown clearly that in order to ensure remote interactions in learning, and to assess changes in the level of knowledge and skills, it is necessary to move from expert teaching methods and assessing results to methods and technologies of a digital environment for communication exchange. Naturally, the educational environment is a huge world with long-term traditions. This article examines the model and means of digitalization of distance learning within the "foreign language acquisition" subject area (on the example of English).

2. BASIS OF ASSESSMENT: BACKGROUND

There is a generally accepted mixed testing technology (oral and written) in the world to determine the level of proficiency in a foreign language. And if the written part of international exams, for example the computer version of TOEFL, can be digital, the oral part is always carried out with the involvement of expert examiners. With the spread of the Corona virus, the movement around the world of experts, who are most often native speakers, has become difficult and sometimes even impossible. Because of this, the dates of international exams were postponed many times. In this regard, the digitalization of the procedure for conducting and assessing speaking skills is especially relevant at the present time [3-4].

The description of the procedures and the assessment methodology in expert form are given in sufficient detail in the methodological support of Cambridge university [5]. According to this generally accepted method, the communicative skills of a person when communicating in a non-native language are assessed by the communication (dialogue) components of real time - speaking and listening. The components of deferred communication - writing and reading - are also checked. At the same time, the conversational components of speaking and writing determine a person's ability to form arbitrary semantics (meaning) by means of a foreign language during direct communication. Listening and reading are reflective activities that determine the ability to recognize the semantics of phrases and the general meaning of a foreign language with the possibility of additional, repeated analysis. Thus, the speaking component is most relevant for determining the methods and building a basic digital model of computer assessment of the target language proficiency level (Table 1).

Table1 Test components

Test components of communicative competence			
Real-time components		Deferred components	
speaking	listening	writing	reading

The assessment of each component has its own characteristics, properties and parameters and should be considered for the purposes of digitalization separately, with the integration of assessments at the final stage of testing.

It should be noted that at present, various on-line tests are quite common to determine the level of English proficiency, for example: EF SET, IELTS, TOEIC, Cambridge English Scale, TOEFL iBT, and others. Some of them are automated, while the correctness of the answer is recorded as a multiple choice: 1 out of N-answers to a written or spoken phrase. This approach has its limitations and is very approximate. More complex testing is done by experts and is quite expensive. In the case of a long-term world pandemic situation, distance assessment of foreign language proficiency should be based on digital technologies, be more accessible, and be quick to modify. Within the scope of this research paper, the digital model is considered only for the speaking component, with the understanding that the proposed principles for constructing digital metrics and assessment are applicable to all test components. In this case, it is assumed that the widespread systems of speech recognition and generation of the "assistant" level are used [6-11].

According to [12,13] the evaluation structure is a level cascade, with a tuple

$$K = \langle H, C2, C1, B2, B1, A2, A1 \rangle$$

where H ... B2 ... A1 are the names of the levels in descending order of the language proficiency level. H - in the document the level without designation. For each alphabetic level, it is proposed to use the following composite fuzzy metrics:

- H, C2, C1:
 Grammatical Resource,
 Lexical Resource,
 Discourse Management,
 Pronunciation
 Interactive Communication

- B2, B1:
 Grammar and Vocabulary
 Discourse Management
 Pronunciation
 Interactive Communication

- A2, A1:
 Grammar and Vocabulary
 Pronunciation
 Interactive Communication

3. PRINCIPLES OF ASSESSMENT METRICS

Below you can see a table of fuzzy metrics for Grammatical Resource and Lexical Resource, combined into Grammar and Vocabulary column (Table 2).

Table 2 Metrics for Grammatical and Lexical resources

	Grammatical Resource	Lexical Resource
	<ul style="list-style-type: none"> Maintains control of a wide range of grammatical forms and uses them with flexibility. 	<ul style="list-style-type: none"> Uses a wide range of appropriate vocabulary with flexibility to give and exchange views on unfamiliar and abstract topics
C2	<ul style="list-style-type: none"> Maintains control of a wide range of grammatical forms. 	<ul style="list-style-type: none"> Uses a wide range of appropriate vocabulary to give and exchange views on unfamiliar and abstract topics.
C1	<ul style="list-style-type: none"> Shows a good degree of control of a range of simple and some complex grammatical forms. 	<ul style="list-style-type: none"> Uses a range of appropriate vocabulary to give and exchange views on familiar and unfamiliar topics.
	Grammar and Vocabulary	
B2	<ul style="list-style-type: none"> Shows a good degree of control of simple grammatical forms, and attempts some complex grammatical forms. Uses appropriate vocabulary to give and exchange views, on a range of familiar topics. 	
B1	<ul style="list-style-type: none"> Shows a good degree of control of simple grammatical forms. Uses a range of appropriate vocabulary when talking about familiar topics. 	
A2	<ul style="list-style-type: none"> Shows sufficient control of simple grammatical forms. Uses appropriate vocabulary to talk about everyday situations. 	
A1	<ul style="list-style-type: none"> Shows only limited control of a few grammatical forms. Uses a vocabulary of isolated words and phrases. 	

The metrics shown in the table are defined as fuzzy. For example, in the metric "Maintains control of a wide range of grammatical forms and uses them with flexibility", it is necessary to define numerically (for the purposes of computer processing) the submetrics: "wide range", "grammatical form", "ease of use". It should be noted that these are the metrics and the submetrics of the 1st level,

since each level from the K tuple also has its own fuzzy gradations and metrics.

3. DIGITALIZATION OF METRICS

3.1 Digital assessment of grammatical forms

Taking into account the general picture, let us consider the metrics of A2 level (Table 3) [14-15]. ...

Table 3 Metrics for A2 level

A2	Grammar and Vocabulary	Pronunciation	Interactive Communication
5	a) Shows a good degree of control of simple grammatical forms. b) Uses a range of appropriate vocabulary when talking about everyday situations.	Is mostly intelligible, and has some control of phonological features at both utterance and word levels.	Maintains simple exchanges. Requires very little prompting and support.
4	<i>Performance shares features of Bands 3 and 5</i>		
3	a) Shows sufficient control of simple grammatical forms. b) Uses appropriate vocabulary to talk about everyday situations.	Is mostly intelligible, despite limited control of phonological features.	Maintains simple exchanges, despite some difficulty. Requires prompting and support.
2	<i>Performance shares features of Bands 1 and 3</i>		
1	a) Shows only limited control of a few grammatical forms. b) Uses a vocabulary of isolated words and phrases.	Has very limited control of phonological features and is often unintelligible.	Has considerable difficulty maintaining simple exchanges. Requires additional prompting and support.
0	<i>Performance below Band 1</i>		

As you can see, fuzzy metrics are also used here, with "overlapping" boundaries. In the combined section "Grammar and Vocabulary" 6 metrics are defined.

In metrics 5a) "Shows a good degree of control of simple grammatical forms", it is necessary to determine the digital forms of fuzzy submetrics "good proficiency" and "simple grammatical forms".

Let us define "a good degree of proficiency" as a certain (small) number of mistakes made in sentences. For example, the number of errors can be correlated with the level $m = \frac{2}{3} = 33\%$ (no more, qualified minority). Here we get a clear digital boundary. If we make a "soft"

boundary, between the levels, having established an indefinite zone between "good" and "bad" language proficiency (for example, from 60% to 75%), when entering it, we use a random choice, or take into account this zone together with the value parameters of other metrics.

Next, we will show the approaches of defining metrics with respect to "simple grammatical forms" and combining them with "good proficiency" metrics.

Simple grammatical forms include:

- verbs: regular, irregular, modal (ability; requests; permission) could (ability; polite requests) would (polite

requests) will (future) shall (suggestion; offer) should (advice) may (possibility) have (got) to (obligation) must (obligation) mustn't (prohibition) need (necessity) needn't (lack of necessity));
 - plural nouns;
 - simple view-temporal forms:
 Present simple:
 Present continuous: present actions and future meaning
 Present perfect simple:
 Past simple: past events
 Past continuous
 Future with going to
 Future with will and shall: offers, promises, predictions, etc;
 - pronouns: personal, possessive and demonstrative;
 - prepositions: place, time, direction and etc.
 In practice, the list of simple forms can be variable, determined by experts and modified at certain intervals of time. Similar conditions are established for experts [16].
 Next, a rectangular matrix is built, where the rows correspond to the above mentioned simple grammatical forms, and the columns correspond to the degrees of proficiency in these forms (Table 4).

Table 4 Rectangular matrix

SGF	Language proficiency (s _n - number of mistakes in %)		
	good	undefined	bad
Regular verbs	m < 25	40 > m >= 25	m > 40
Irregular verbs	m = < 25	-	m > 25
...			
Past continuous
...	...		
Personal pronouns	m < 30	-	m >= 30

Thus, you can get differential digital data for metric 5a). The integral value can be obtained as an average for all SGFs:

$$M [5a)] = (\sum s_n) / N;$$

or as a weighted average

$$M [5a)] = (\sum k_n * s_n) / N;$$

More detailed personalized assessment can be done by constructing a histogram according to the data in Table 4. This histogram, similar to psychological tests, allows you to form a personal digital profile of the student's knowledge level [17-18].

3.2 Digital assessment of vocabulary usage

Next, consider the possible metrics for the condition A2: 5b) -"Uses a range of appropriate vocabulary when talking about everyday situations". Here are two fuzzy descriptions:

- "everyday";
- situations ";
- "range of appropriate vocabulary".

To construct appropriate digitalization metrics, it is necessary at least to transform the fuzzy definitions into one set of elements with the possibility of digital parameterization. According to [19], situation is the set of things that are happening and the conditions that exist at a particular time and place.

Using this definition, let us also pay attention to the fact that in many modern textbooks, for example [Oxford, Cambridge etc] new words, grammatical and lexical elements, rules, exceptions, etc. are considered in the format of description for a certain situation related to a specific topic. A new fragment (component) of student's knowledge (Δ k) is associated with the components of the presentation situation. The components of the situation are words and other lexical units related to the topic and the level of students.

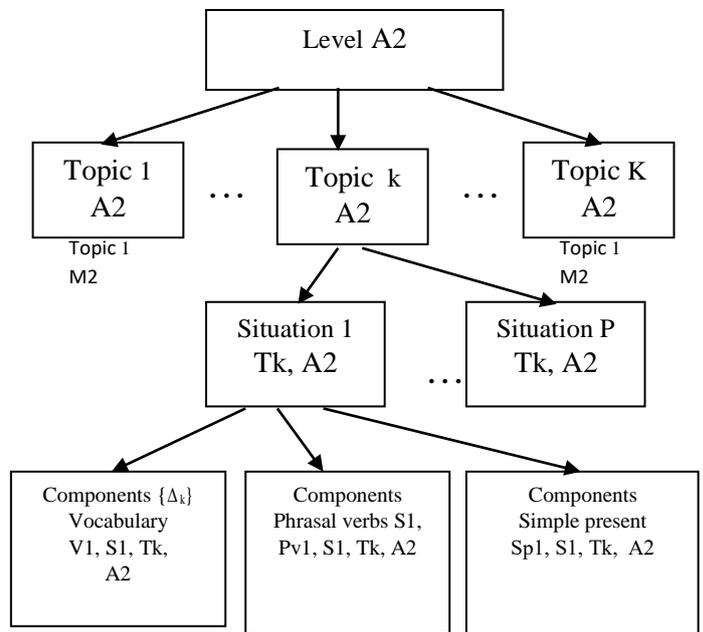


Figure 1 The digital model structure of A2 knowledge components nesting

A graphical representation of the digital model of condition A2: 5b) is shown in Fig. 1.

Any knowledge component (Δ) included in the structure of Fig. 1 is an element or a combination of elements included in the hierarchy of a tuple of sets:

$$\Delta K \in \langle CS, SN, TK, LL \rangle,$$

which allows using such a representation for digital computer processing.

It is obvious that conducting tests and exams other situations can be used, but comparable to the previous one related to a specific topic, using a certain set of words and typical patterns studied in this topic. There are usually 15-25 such topics, for example, in KET[. Each topic includes a list (several dozen) of words for study. That is, we can say that testing the knowledge and skills obtained during the study of a course of a certain level should correspond

to the level and be carried out within the framework of the topics set proposed for the exam (for example: study, attending classes, going to the cinema, breakfast, transport, entertainment, etc.), respectively, everyday situations within the framework of these topics and sets of learned words related to them.

The topic of communicative situations is set based on the requirements of the course being studied.

For example, for the topic «education» A2, the following situations can be proposed:

- describe your background education;
- talk about your favorite subjects and teachers;
- describe a typical school day;
- describe the class or laboratory in which the classes took place;
- preparation for tests and exams;
- getting books in the library;
- find out from classmates the schedule for tomorrow, etc.

Thus, instead of enumerating all the joint combinations of testing conditions: topics, topic situations, topic words, lexical level elements, we get a decomposition of these conditions, which can be formed in the format of linked database tables, where data means string variables, grammatical patterns, names topics, names of situations and the words themselves.

3.3 Digital assessment of lexical forms: Phrasal verbs

Phrasal verbs in English are syntactic two or three-word constructions that include a verb supplemented by a preposition or adverb, as well as an adverb and preposition [19]. In addition, other classifications are used: syntactic (separable / inseparable) and semantic (transitive / intransitive), but this is not the focus at the A2 level. Below you can see the tables 2,3 [20], with examples of such constructions, showing the change in the semantics of the corresponding verb, depending on the complement. In these examples, verbs define an action, and complements define the direction of that action.

Table 2. The verb "come"

come	<i>up</i>	- повышаться - возникнуть, появиться
	<i>down</i>	- спуститься вниз - упасть
	<i>in</i>	- войти
	<i>out</i>	- выйти из - выйти (о книге) - стать известным (об информации, секретах) - закончиться каким то результатом
	<i>back</i>	- вернуться

Table 3. The verb "look"

look	<i>up</i>	- смотреть вверх - посмотреть (поискать) какую-то информацию в справочнике
	<i>down</i>	- смотреть вниз, опустить взгляд
	<i>in</i>	- посмотреть, заглянуть
	<i>out off</i>	- выглянуть
	<i>back</i>	- оглянуться
	<i>away</i>	- отвернуться
	<i>for</i>	- искать, надеяться
	<i>after</i>	- присматривать за кем-то, заботиться, ухаживать

The syntax of the construction of a phrasal verb is quite simple and allows you to determine its application by recognizing the verb and the subsequent presence of complement.

Such a bundle, minimally burdened with additional conditions and restrictions, is very simply represented in the database as "one-to-many". Additional conditions include, for example, the definition of the tense of the verb. However, the presence of a base with all tense forms of verbs allows you to form a digital metric for defining the primary verb.

The main difficulty of assessment here is that it is necessary to take into account the following semantic issues:

- the appropriateness of the phrasal verb use;
- determination of the action direction correctness (intention).

The metric of the validity of the application can be determined based on the topic of communication. If the applied verb is included in the vocabulary set of one of the given topics and lower levels, then the appropriateness of its use is true. Otherwise, an error is recorded.

It is more difficult to form the directional metric due to its semantic basis. If we consider only the statement - the answer, then there is uncertainty about the direction of the action (intention). To eliminate this ambiguity, it is necessary to form the statement in such a way that the direction of the action (intention), expressed by the phrasal verb used in the answer, is uniquely determined by the given phrase. In this case, the metric of absolute directivity in space can be determined by three-dimensional Cartesian coordinates with a zero point (Fig. 2), and the directivity relative to the subject (object) is determined the direction of the perpendicular to the surface of the sphere with the subject (object) in the center. The time metric is formed similarly on the one-dimensional axis (Fig. 2, 3).

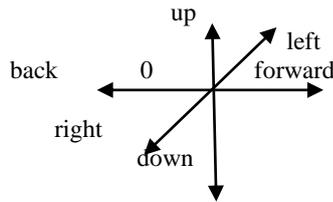


Figure 2 Three-dimensional Cartesian axes

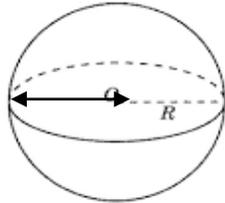


Figure 3 Sphere of directions

In the first case, we have the possibility of digitizing not only in the direct directions of the axes, but also in the intermediate directions.

In the second, it becomes possible to digitize permanently in the relative direction of action - to the center (object) / from the center (from the object). And also, if necessary, remoteness from the center (for example, using several spheres of different diameters).

Similar digital metrics can be generated in other cases, for example, to determine the correctness of the prepositions in, on, at, and others usage.

4. CONCLUSIONS

Modern methods of testing and assessing the level of foreign language proficiency are face-to-face and time-consuming, and costly to organize. As part of the digitalization of educational systems and services, it is necessary to change assessment technologies switching from full-time to distance formats. Modern methods of speech recognition do not allow at the present time to carry out reliable assessment of the level of language proficiency.

The approach proposed in the paper is based on the structuring of the method of full-time expert assessment. This allows us to propose the creation on this basis of a digital model and digital metrics for assessing the level of foreign language proficiency. Speech recognition at the level of well-known voice assistants is enough.

In this case, it is also possible to use traditional computer technologies that do not require significant resources for information processing.

REFERENCES

[1] Teleworking during the COVID-19 pandemic and beyond. A practical guide, Geneva: International Labour Office, July 2020.

[2] Savić Dobrica, COVID-19 and Work from Home: Digital Transformation of the Workforce, May 2020. Grey Journal 16(2), pp.101-104.

[3] Digital Transformation of Learning. web: <https://www.valamis.com/hub/digital-transformation-of-learning>.

[4] N.A. Kameneva, L.V. Zenina, E-learning-for-developing-knowledge-and-skills-in-english-language-teaching. Journal "Economics, Statistics, and Informatics 2013, issue 5, pp. 3-6.

[5] Cambridge English: Key handbook for teachers, Cambridge University Press, 2018, P. 14.

[6] Startling digital divides in distance learning emerge. web: <https://en.unesco.org/news/startling-digital-divides-distance-learning-emerge>.

[7] Superfast speech recognition without servers using a real example. Usage mode: <https://habr.com/ru/post/237589/>.

[8] E.V. Polyakov, et al. Development of an intelligent voice assistant and study of the learning ability of natural language recognition algorithms.

[9] Ayn de Jesus. AI for Speech Recognition – Current Companies, Technology, and Trends.

[10] Speech Recognition Technology Overview.

[11] Pierrick Milhorat (Institut Mines-Télécom) a.o. - Building the Next Generation of Personal Digital Assistants

[12] Nicholas Fearn, Brian Turner.- Best text to speech software of 2020: Free, paid and online voice recognition apps.

[13] <https://buyingbusinessstravel.com.ru/news/technology/41740-zashchitnaya-mask-perevodchik-poyavilas-v-yaponii/>

[14] <https://www.cambridgeenglish.org/exams-and-tests/cefr/>

[15] <https://www.coe.int/ru/web/lang-migrants/cefr-and-profiles>

[16] L.Yu. Shchipitsina, Information technology in linguistics: textbook. Manual, L.Yu. Shchipitsin, FLINT: Nauka, 2013, 128 p.

[17] Information technologies in linguistics: a textbook for students of institutions of higher professional education, A.V. Zubov, A.I. Zubova, 2nd ed. Rev, Publishing Center "Academy", 2012 p.

[18] A.N. Malyavina, Inform technologies in linguistics: Textbook Manual, Togliatti publishing house of TSU, 2013, 80 p.