A Multimedia Corpus for Language Teaching Purposes: the *MagyarOK Video Corpus*

Katalin Pelcz

International Studies Center, University of Pécs Medical School
(<pelcz.kata@pte.hu>)

Abstract

This article presents the structure and key elements of the *MagyarOK videókorpusz* (*MagyarOK* video corpus). The content of the video corpus is based on and goes beyond the *MagyarOK* textbook series that takes learners from the A1 to the B2 level of the CEFR. This topic-centered pedagogical corpus was created with the specific aim of making variability and repetition of topic-related vocabulary observable to language learners. The corpus is rather small and contains two types of data: narrative texts and dialogues, both illustrating the mentioned topic-related lexical repetitions and variations. The text types illustrate the characteristics of natural language use in particular contexts and situations. The article argues that being able to observe such texts should be an organic part of learning right from the outset. The discussed data were examined both quantitatively and qualitatively using tools for corpus linguistics and discourse analysis. The article also describes the video corpus itself and some of its possible uses.

Keywords

corpus pedagogy; Hungarian as a foreign language; language learning; topic-centered pedagogical corpus; video corpora

1. Introduction

This article presents the *MagyarOK video corpus*, a special corpus for pedagogical (language teaching) purposes. This corpus is freely available as

a learning tool for language learners[1]. The video corpus is an integral part of the larger *MagyarOK pedagógiai korpusz* (*MagyarOK* pedagogical corpus), a database that comprises various text types and multimedia materials to support language learning. The video corpus aims at integrating didactic language elements into the learning process right from the outset such as:

—   Lexico-grammatical features linked to the theme of the recording,
—   Pragmatic features,
—   Materials to observe and practice pronunciation and prosody,
—   The possibility to observe frequently used linguistic elements associated with the discussed topic.

In short, this video corpus has been built with the specific purpose to become a methodological tool for language teaching and learning that offers numerous opportunities for observation and memorisation, pronunciation training as well as means to develop the students' linguistic intuition or their «feel for the language» (Gabrielatos 2005). The corpus is structured by proficiency level and topic.

2. Designing criteria

The *MagyarOK* corpus was designed with the aim of complementing the *MagyarOK A1-B2* textbook series for Hungarian as a foreign language (Szita, Pelcz 2013-2019). It is a pedagogical corpus with the primary aim of making language teaching and learning more effective (cf. Szita 2021), which brings together the texts that learners encounter in the process of learning (cf. Aston 2001; Willis 2003; Braun 2006). The *MagyarOK video corpus* is a subcorpus of this collection freely accessible on the Sketch Engine website for corpus analysis. Advantages of pedagogical corpora such as the *MagyarOK video corpus* are manifold; learners are able to recontextualise the recorded communicative situations (Widdowson 2003) as they rely on the material covered in their course books. They can also see interactions realised outside the classroom and reevaluate them from various perspecti-

---

[1] <https://www.youtube.com/channel/UCv4OI7fdvsdcXhe8DPyyQrg> (11/2022).

ves (focus on pronunciation, intonation, lexical units, etc.), and bring this new knowledge back to the classroom by reshaping it according to their needs. Furthermore, pedagogical corpora in general can also contribute to independent language learning as it enables students to discover certain regularities and useful linguistic elements without the help of the teacher (Szirmai 2005, 35).

The *MagyarOK video corpus* is designed according to the following criteria:

— The video corpus does not aim to be representative. It is designed for lower proficiency levels with the purpose of providing them with a collection of accessible texts. As the videos present similar situations at more than one level, they also allow for step-by-step linguistic progression.
— The recordings are adapted to the learner's proficiency level, but at the same time go slightly beyond the course book texts and provide, thus, a larger input for natural-sounding language use.
— Learners can therefore integrate the observed vocabulary and structures into their linguistic products.
— At lower proficiency levels (A1-B1), topics are repeated and expanded in a circular manner. In the case of descriptive texts, several texts with similar content are presented, which makes frequently used topic-related elements salient; these elements are helpful to language learners for successful communication. In case of dialogues, the actors improvised based on a loosely constructed script. The accent lies here on the intertwin of topic-related and conversational elements of language so that these dialogues provide opportunities for students to observe the dynamics of discourse and integrate certain elements into their own language use.

3. Working with the video corpus: data and examples

In the first two decades of the 21st century, the use of videos in language teaching has still been one of the complementary materials. The offer of the major publishing houses was characterised by the triad of textbook/

workbook/CD, with occasional video recordings on the websites of the textbook series. These videos were mostly produced for independent use outside the classroom, intended by the authors and used by students and teachers. The most obvious way to develop interpersonal communication was not through short films for textbooks, but through various types of interactions in the classroom.

As mentioned above, the video corpus complements and expands the content of the *MagyarOK* textbook series. Working with the videos is therefore preceded by introducing the lexical and grammatical components of the topic to the student. This is done with the help of the textbook. Then, the student encounters the same topic in the form of a monologue in the video corpus. This video presentation goes slightly beyond the written text from the lexical point of view but the fact that language perception is greatly aided when the language learner can see the speaker/speakers, the communicative situation, contributes to a better understanding of what is being said. Several variations on the same topic follow this first recording that provide space for extended observation and memorization of commonly occurring items.

According to the Common European Framework of Reference for Languages (CEFR), learners typically reach entry level B1, which is the level at which they can become independent language users. We add to this statement that when learners reach the level of experience and linguistic competence to participate as equal partners in a communicative situation depends largely on individual factors: the level of language proficiency and the learner's personality (Oxford, Ehrman 1992). Frequent encounters and variations with similar linguistic items can speed up language acquisition on the one hand and increase the self-confidence of language learners on the other. Videos develop language flexibility. Cognitive flexibility and control play an important role in the learning process, enabling context-appropriate (linguistic) behaviour and selective reception and processing of new information (Egner, Hirsch 2005). Cognitive control excludes distracting information by monitoring language experiences and incorporates frequently experienced content, thus defining it as relevant. This complex system enables us to adapt to the communicative situation (cf. Cohen 2017). Thus, presenting learners with linguistic content in a clear

and unambiguous way across multiple text types and registers, contributes to the organic development of the learner's language use.

The videos for the *MagyarOK* series can be found on the YouTube channel *Magyarul beszélni menő!* (*Speaking Hungarian is cool!*)[2].

The channel contains 143 videos distributed in the following way:

— 62 videos for the A1 level,
— 47 videos for the A2 level,
— 34 videos for the B1 level.

The videos are generally between half a minute and a minute long, the shortest being 11 seconds and the longest 2,59 minutes. The sub-corpus of the video transcripts contains 8856 tokens. The videos are included in four playlists, grouped by language level. There are two playlists for the A1 level, where the videos for the textbook and a complementary reading book are available separately. The transcripts are available on the Sketch Engine website where they can be analyzed with corpus tools[3].

The design of the videos is minimalist, so that students at lower language levels are not distracted by different levels of meaning (environment, objects, etc.), but can focus on the text, its content and form, and the gestures and mimics of the speakers. The recordings were made with the help of actors from the Janus University Theatre in Pécs. The actors first cited the text in the course book (this text was based on the analysis of authentic texts), then freely adapted it to create similar narrative texts, and, eventually, to present dialogues to the theme. The videos therefore contain the texts of the textbook as well as their paraphrases. The videos are available with or without subtitles in Hungarian.

Let us look at two examples to Chapters 7 (Szita, Pelcz 2022, 95) and Chapter 9 (ivi, 119) of the *MagyarOK* A1+ textbook. In the first example, the textbook text of Chapter 7 is about the eating habits of Hungarians. At this point, the students have attended about 60 Hungarian lessons.

---

[2] <https://www.youtube.com/channel/UCv4OI7fdvsdcXhe8DPyyQrg> (11/2022).
[3] <https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fmagyarok> (11/2022).
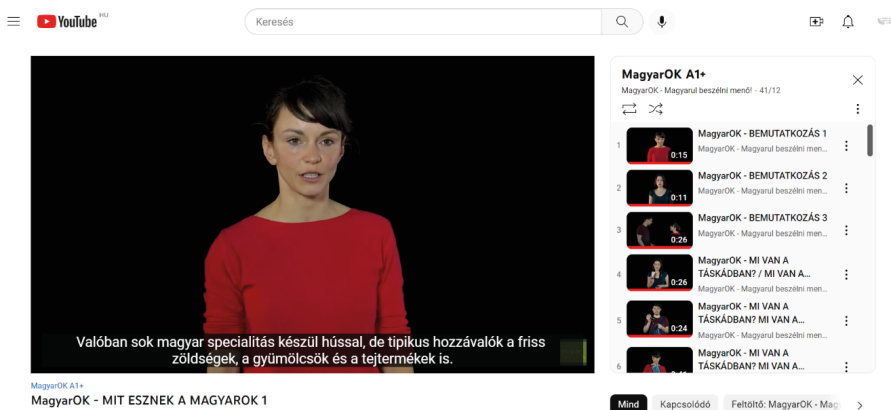
1. Mit esznek a magyarok?

Közhely, hogy a magyarok húst hússal esznek. Valóban sok magyar specialitás készül hússal, de tipikus hozzávalók a friss zöldségek, a gyümölcsök és a tejtermékek is. Ismert magyar ételek a pörkölt, a gulyás, a halászlé, a töltött káposzta, a paprikás csirke, a bableves és a túrós csusza. A magyar konyha elég kevés fűszert használ. Tipikus fűszerek a só, a bors és a pirospaprika.

A magyarok általában nem sok időt töltenek az asztalnál. A főétkezés a meleg ebéd. Ilyenkor majdnem mindig van leves is. Sajnos a magyar ételek nem mindig egészségesek. Sokszor túl fűszeresek, zsírosak és nehezek, de nagyon finomak.

What do Hungarians eat?

It is a cliché that Hungarians «eat meat with meat». Many Hungarian specialties are indeed made with meat, but fresh vegetables, fruit and dairy products are also typical ingredients. Famous Hungarian dishes include stew, goulash, fish soup, stuffed cabbage, chicken with paprika, bean soup and cottage cheese. Hungarian cuisine uses very few spices. Typical spices are salt, pepper and red pepper.

Hungarians generally do not spend much time at the table. The main meal is a warm lunch. There is almost always soup at this time. Unfortunately, Hungarian food is not always healthy. It is often too spicy, fatty, and heavy, but very tasty.

Table 1 – Course book text about Hungarian eating habits read by the actress. The recording can be viewed with or without subtitles.

Another video recording complements the original text, which is a short dialogue presenting the key information of the course book text in the form of questions and answers. This second video does not necessary expand students' knowledge on the theme, but it opens the door to a wide range of observable features related to interactive texts such as discourse co-construction, discourse markers, non-verbal elements of communication, etc. This step-by-step procedure – first a text already known to the student presented as a monologue, then a dialogue to expand learners' toolkit for interactive situations and then, in case of other texts, further variations – gradually adds new linguistic knowledge to already acquired elements while giving learners the opportunity to see how the new items are used in communication. The procedure combines several hypotheses of Stephen Krashen's Theory of Second Language Acquisition (1988): the Acquisition-Learning hypothesis, the Monitor hypothesis, and the Input hypothesis. Krashen distinguishes between an «acquired system» and a «learned system» in the language learning process. The multi-stage incorporation of videos into the language learning process links acquisition and learning, and the Monitor Hypothesis highlights the relationship between the two, showing how learning affects acquisition. The Input Hypothesis states that learners progress effectively in language learning when they receive a large linguistic input that goes one step beyond their current level of linguistic competence.

The transcription and the translation of the improvised dialogue reads as follows:

2.
- Milyen a magyar konyha?
- Közhely, hogy a magyarok húst hússal esznek.
- A magyarok nem is esznek zöldséget?
- De, azért tipikusan eszünk friss zöldségeket, gyümölcsöket és gyakori a tejtermék is.
- Milyen ismert magyar ételek vannak?
- Hát, ott van például a pörkölt, vagy a leghíresebb talán a gulyás vagy a halászlé, a töltött káposzta, van a paprikás csirke vagy a bableves, esetleg a Túró Rudi, az ilyen tipikusan magyar étel, illetve a túrós csusza.

- A magyar konyha sokféle fűszert használ?
- Nem, alapvetően kevés fűszert használunk, ellenben nagyon sokat belőle. Tehát a magyar ételek fűszeresek, és nagyon zsírosak szoktak lenni, de ellenben nagyon finomak.
- Tehát a magyar konyha egészséges?
- Nem, de annál finomabb.

- What is Hungarian cuisine like?
- It is a cliché that Hungarians eat meat with meat.
- Don't Hungarians eat vegetables?
- Yes, we typically eat fresh vegetables and fruit, and dairy products are also common.
- What are some well-known Hungarian dishes?
- Well, there's stew, for example, or perhaps the most famous is goulash or fish soup, stuffed cabbage, paprika chicken or bean soup, or Túró Rudi, which is also a typically Hungarian dish, or túrós csusza.
- Does Hungarian cuisine use a wide variety of spices?
- No, in fact, we use very few spices, but we use a lot of them. So Hungarian food is spicy and tends to be very fatty, but it's very tasty.
- So Hungarian cuisine is healthy?
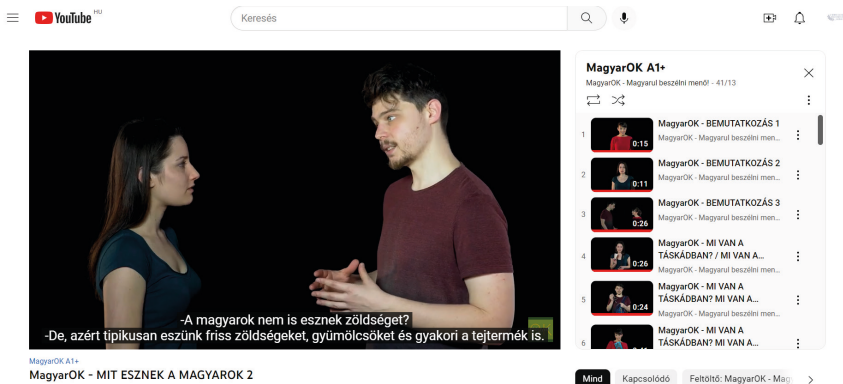*- No, but it is really tasty.*

Table 2 – Actors' improvisation based on the text about Hungarian eating habits. The recording can be viewed with or without subtitles.

4. Thematic processing of videos for vocabulary acquisition

Let us now compare the two texts. The underlined expressions in bold are typical elements of spoken language that only emerged in the dialogue but were not part of the original course book text.

1. *Mit esznek a magyarok?*	2. - *Milyen a magyar konyha?*
Közhely, hogy a magyarok húst hússal esznek. Valóban sok magyar specialitás készül hússal, de tipikus hozzávalók a friss zöldségek, a gyümölcsök és a tejtermékek is.	- *Közhely, hogy a magyarok húst hússal esznek.* - *A magyarok __nem is__ esznek zöldséget?* - *__De, azért__ tipikusan eszünk friss zöldségeket, gyümölcsöket és gyakori a tejtermék is.* - *Milyen ismert magyar ételek vannak?*
Ismert magyar ételek a pörkölt, a gulyás, a halászlé, a töltött káposzta, a paprikás csirke, a bableves és a túrós csusza.	- *__Hát, ott van például__ a pörkölt, vagy a leghíresebb __talán__ a gulyás vagy a halászlé, a töltött káposzta, van a paprikás csirke vagy a bableves, __esetleg__ a Túró Rudi, __az ilyen__ tipikusan magyar étel, illetve a túrós csusza.* - *A magyar konyha sokféle fűszert használ?* - *Nem, __alapvetően__ kevés fűszert használunk, ellenben nagyon __sokat belőle__. __Tehát__ a magyar ételek fűszeresek, és nagyon zsírosak szoktak lenni, de ellenben nagyon finomak.*
A magyar konyha elég kevés fűszert használ. Tipikus fűszerek a só, a bors és a pirospaprika.	- *__Tehát__ a magyar konyha egészséges?* - *Nem, __de annál finomabb__.*
A magyarok általában nem sok időt töltenek az asztalnál. A főétkezés a meleg ebéd. Ilyenkor majdnem mindig van leves is. Sajnos a magyar ételek nem mindig egészségesek. Sokszor túl fűszeresek, zsírosak és nehezek, de nagyon finomak.	

Table 3 – The original text from the *MagyarOK* A1 course book (left column) and the transcription of the improvised dialogue (right column). Spontaneously emerging elements of the spoken language are underlined.

The improvisation contains a fair amount of interactional language, and the actors make use of numerous common conversational strategies in a natural way. For example, they repeat their partner's words (- *Does Hungarian cuisine use a wide variety of spices? - No, in fact, we use very few spices, but…),* draws conclusions based on text coherence (*So Hungarian food is spicy and tends to be very fatty, but it's very tasty. - So Hungarian cuisine*

is healthy?), hesitate (- *Well, there's stew, for example, or…*) or give positive feedback on what has been said (- *So Hungarian cuisine is healthy? - No, but it is really tasty).* These frequently used simple strategies are easy to follow at lower proficiency levels and learners can integrate them into their conversations to bond with their partner(s).

As the above examples show, the *MagyarOK* videos show everyday narrative texts on the one hand and improvised, non-scripted conversations to the given topic. The videos also allow us to observe important elements of non-verbal communication, which can emphasise, contradict to, replace or complement verbal expressions (Róka 2002). This is a particularly important task, since in our everyday conversations a small percentage of non-verbal communication is vocal, which is outweighed by metacommunicative components (cf. Mehrabian 1981). Although the quoted theory has been disputed by many, it does highlight the fact that a significant percentage of communication takes place at the non-verbal level.

A major advantage of videos over audio recordings is that the visual input helps students to rate their comprehension higher (Coniam 2001). This is not only because of the non-verbal means or the visible movements of the actors' lips, but also because the content of the texts is familiar and challenging at the same time. The frequent repetitions in the text variations facilitate vocabulary retention and, thus, the incorporation of these linguistic items into the use of the target language.

The following example illustrates repetition and expansion in narrative texts. We can observe how grammatical forms are repeated and consolidated, and how students are given bits of natural language to adapt the topic to their own content while observing some tendencies of natural language use. The example is taken from Chapter 9 of the *MagyarOK* A1+ textbook, the topic is «Hobbies and objects».

The short texts were improvised by the actors. Working with actors had many advantages. Firstly, they did not know the content of the syllabus, so they relied on their own judgement to decide what language was 'simple enough'. They were given a sentence and were free to make up content by association, using it as a model. The actors do not have any experience as language teachers, which means that they do not represent a filter to adapt the language product to the students' level of knowledge. They were not

inclined to create unnatural, easily understood but atypical utterances. They understood immediately what was expected of them and could improvise easily in front of the camera.

First, the four short texts of the course book are recited (texts 1, 3, 5 and 7), they are then followed by a variant as the examples show. The repetitions and expansions are marked with colours.

1. *Van egy régi gitárom. Általában a barátommal gyakorlunk. Rockzenét játszunk. Néha koncertezünk is.*	1. *I have an old guitar. I usually practice with my friend. We play rock music. Sometimes we play gigs.*
2. *Van egy régi túrabakancsom. Általában a testvéremmel túrázunk. Hegyet mászunk, néha erdőbe is megyünk.*	2. *I have an old hiking boot. I usually go hiking with my brother. We climb mountains, sometimes we go into the woods.*
3. *Nagyon szeretek sportolni. Ha van egy kis szabadidőm, pingpongozom a barátnőmmel. Van egy pingpongasztalom a kertben, ott játszunk. A barátnőm nagyon jól tud pingpongozni.*	3. *I really like sports. When I have some free time, I play ping-pong with my girlfriend. I have a ping pong table in the garden, we play there. My girlfriend is very good at ping pong.*
4. *Nagyon szeretek ukulelézni. Van egy ukulelém. Ha van egy kis időm, mindig játszom rajta. A barátnőmnek van egy gitárja. Néha együtt zenélünk.*	4. *I really like playing the ukulele. I have a ukulele. If I have some time, I always play it. My girlfriend has a guitar. Sometimes we play music together.*
5. *Ha van egy kis szabadidőm, zongorázom. Van egy régi zongorám, azon gyakorolok. Általában jazzt vagy komolyzenét játszom.*	5. *When I have some free time, I play the piano. I have an old piano, I practice on it. I usually play jazz or classical music.*
6. *Ha van egy kis szabadidőm, énekelek. A barátnőmmel gyakran éneklünk duettet. Előfordul, hogy koncerten is fellépünk.*	6. *If I have some spare time, I sing. My girlfriend and I often sing duets. Sometimes we also perform in concerts.*
7. *Ha van egy kis időm, természetesen számítógépezek. Van egy új gépem. Sok mindent lehet vele csinálni: játszani, programozni, internetezni, filmeket nézni.*	7. *When I have some spare time, I play on the computer, of course. I have a new computer. I can do many things with it: play games, program, surf the Internet, watch movies.*
8. *Ha van egy kis időm, természetesen gördeszkázom. Van egy új deszkám. Sok mindent lehet vele csinálni: trükköket tanulni, de csak közlekedni is.*	8. *When I have a little time, of course I skateboard. I have a new skateboard. I can do a lot of things on it: learn tricks, but also just get around.*

Table 4 – Short model texts
(texts from the course book and improvised variations) about hobbies.

5. Thematic processing of videos from a grammatical perspective

Watching these short videos help students to memorise certain lexical items, such as: «When I have some free time, ... When I have some time, ...», whose correct pronunciation and intonation can be observed across several texts. Learners can also see enough linguistic variation to find those that fulfill their needs when they wish to present their own hobbies.

In addition to recording lexico-grammatical elements and collocations, videos can be used for a wide range of purposes. Not only do they provide opportunities for the observation of longer texts with similar (but not identical) informational content, but variations on the same theme also allow the observation of a number of grammatical elements. Note the above example from Chapter 9 of *MagyarOK* A1+ (Szita, Pelcz 2022, 121). These repetitions and variations can be presented as 'concordance lines', that is, words with similar functions can be listed one below the other and marked with different colours. This manner of sentence presentation enables learners to observe tendencies and rules of Hungarian word order.

Van Milyen? Mim?	*have What kind of? What?*
Van egy régi zongorám, azon gyakorolok.	I *have an old piano I practice on.*
Van egy régi túrabakancsom.	I *have a pair of old hiking boots.*
Van egy új gépem.	I *have a new camera.*
Van egy új deszkám.	I *have a new board.*

Table 5 – Similar structures in more than one text presented as concordance lines.

After the course book texts and related text variants have been discussed, the final task for the students is to create their own text using the many examples provided in the recorded texts. Thus, the texts serve as models for the students' own linguistic productions to enhance their linguistic quality and naturalness.

The transcriptions of the video corpus can also provide valuable opportunities to observe selected linguistic points. These points can be analyzed with corpus methods, for which an example is provided here. The lemma *szeret* 'to like' was taken as an illustrative example because this is the most commonly used verb in the corpus and, of the conjugated forms, the singular forms are the most frequently used. The present tense, singular first person,

definite or indefinite[4], declarative or conditional verb forms *szeretek* 'I like' (indefinite conjugation), *szeretem* 'I like' (definite conjugation), *szeretnék* 'I would like' (indefinite conjugation), *szeretném* 'I would like' (definite conjugation) are frequently used in informal everyday language as speakers often discuss their preferences, wishes and desires. The following table shows the four conjugated forms of this verb as they appear in the corpus of the video transcriptions.

This presentation reveals that each grammatical form typically appears in relation to certain topics. For example, the conditional mode (*szeretnék, szeretném*) does not appear in the corpus for the topic *Work*, and that the indefinite conjugation (*szeretek, szeretnék*) is not used for the topic *People*. Naturally, this does not mean that expert language users could not form utterances related to any of the listed topics with any of the verbs, but the video corpus tells us that certain structures are more closely associated with the topics. Similarly, it is interesting to observe, that (1) negation can be linked to any of the forms, (2) the most frequent modifier of the verb *szeret* 'like' is *nagyon* 'very', (3) and the verb is usually placed at the beginning of the sentence to receive emphasis[5].

Such and similar observations can help topic-focused processing starting from a grammar point and contribute, in this manner, to consolidating linguistic knowledge and to developing linguistic intuition and flexibility. Let us look at some typical examples from the corpus sorted by verb and topic (Table 6), the numbers indicate the number of times a given verb occurs in the video corpus.

---

[4] A characteristic feature of the Hungarian language is the use of the indefinite and definite verbal conjugations. Intransitive verbs only have an indefinite conjugation while transitive verbs have both conjugations. The conjugation of the verb is determined by type of the direct object that can also be definite or indefinite. A definite direct object requires the definite conjugation, and an indefinite direct object is followed by the definite conjugation.

[5] Hungarian has a flexible word order. This means that speakers have considerable freedom to decide which part of the sentence should be emphasised and can build their utterances accordingly.

szeretek: declarative mood, present tense, indefinite conjugation, singular first person	**szeretem:** declarative mood, present tense, definite conjugation, singular first person	**szeretnék:** conditional mood, present tense, indefinite conjugation, singular first person	**szeretném:** conditional mood, present tense, definite conjugation, singular first person
Lakás, élettér Housing, living space **7 results**	**Lakás, élettér** Housing, living space **4 results**	**Lakás, élettér** Housing, living space **3 results**	**Lakás, élettér** Housing, living space **0 results**
Miért **szeretek** nagyvárosban *lakni*? **'Why do I like living** in a big city?' Nem **szeretek** itt *lakni*. Túl nagy ez a város. **'I don't like living** here. The city is too big.' Azért **szeretek** itt *lakni*, mert bármikor ki tudok menni a kertbe és végre van egy labradorom is. **'I like living** here because I can go out in the garden anytime and I finally have a Labrador.' Nekem is van egy kis erkélyem, de nem nagyon **szeretek** *kimenni*, mert nagyon forgalmas helyen van. 'I also have a little balcony, but **I don't really like to go out** because it's very busy.'	Nagyon **szeretem** *Japánt*, gyönyörű ország. **'I really like/ love Japan**, it's a beautiful country.' Gyermekkoromban sokat voltunk ott, és nagyon szerettem ott a Dunában fürdeni, fára másztunk, filmet forgattunk… úgyhogy nagyon **szeretem** *a Duna-partot*, egyszer talán veszek ott egy házat. 'When I was a kid, we used to go there a lot, and I loved swimming in the Danube there, climbing trees, making movies... so I **love the banks of the Danube**, maybe one day I'll buy a house there.'	Csendesebb helyre **szeretnék** *költözni*. **'I want to move to a quieter place.'** Olyan helyen **szeretnék** *lakást bérelni*, ahol jobb a közlekedés. **'I want to rent an apartment** in a place with better transport connections.' Én nem **szeretnék** *egyedül maradni* egy nagy családi házban. **'I don't want to be alone** in a big family house.'	
Nyelvtanulás Language learning **0 result**	**Nyelvtanulás** Language learning **5 results**	**Nyelvtanulás** Language learning **3 results**	**Nyelvtanulás** Language learning **1 result**

	Délelőtt japánórák vannak. Ezek nagyon érdekesek. **Szeretem** *ezt a nyelvet*. Már sok mindent értek, de olvasni és írni nagyon nehéz. '*There are Japanese lessons in the morning. They are very interesting.* **I like this language**. *I understand a lot now but reading and writing is very difficult.*' Délelőtt németórák vannak. Ezek nagyon érdekesek. **Szeretem** *ezt a nyelvet*, csak nehéz a kiejtés. '*In the morning there are German lessons. They are very interesting.* **I like this language**, *but the pronunciation is hard difficult.*' **Szeretem** *a latin nyelveket*, a francia különösen tetszik. '**I like Latin languages**, *French in particular.*'	**Szeretnék** *megtanulni olaszul*, mert nagyon tetszik a nyelv. '**I** *want to learn Italian because I really like the language.*' Ő jól beszél olaszul, és én **szeretnék** *megtanulni* legalább egy alapszinten *olaszul* minél előbb. '*She/He speaks Italian well and* **I would like to learn** *at least a basic level of* **Italian** *as soon as possible.*' Azért **szeretnék** *olaszul megtanulni*, mert a barátom kint él Olaszországban. '**I would like to learn Italian** *because my boyfriend lives in Italy.*'	Én csak **szeretném** egy kicsit *elmélyíteni a tudásomat*. Tetszik a nyelv. '**I** *just* **want to deepen my knowledge** *a bit. I like the language.*'
Emberek, élőlények **People, living beings** **0 result**	**Emberek, élőlények** **People, living beings** **5 results**	**Emberek, élőlények** **People, living beings** **0 result**	**Emberek, élőlények** **People, living beings** **4 results**
	Nagyon **szeretem** a családomat. '**I love** *my family* **very much**.' Nagyon **szeretem** a kutyákat. '*I love dogs very much.*' …**Szeretem**, mert nagyon kedvesek az emberek. '…**I love them** *because they are very nice people.*'		**Szeretném** *bemutatni Zolit*, a barátomat. '**I would like to introduce Zoli**, *my friend.*' **…szeretném** kicsit *bemutatni a pályáját*. '…**I would like to** *briefly* **present his career**.'
Munka, szabadidő **Work, free time** **8 results**	**Munka, szabadidő** **Work, free time** **4 results**	**Munka, szabadidő** **Work, free time** **0 result**	**Munka, szabadidő** **Work, free time** **0 result**

Nagyon **szeretek** *sportolni*. '**I really like sports.**' Nagyon **szeretek** *ukulelézni*.' '**I really like playing the ukulele**.' Nagyon **szeretek** *Sárával lenni*, mert intelligens, türelmes és kedves. '**I really like being with Sára** because she is intelligent, patient and kind.' Nagyon **szeretek** *tanítani*. 'I really like teaching.' Nagyon **szeretek** *festeni*, már volt néhány kiállításom is. '**I like painting very much.** I have had some exhibitions.'	**Szeretem** *a munkámat*, mert nagyon érdekes és változatos. 'I like my job because it is very interesting and varied.' *A filmeket* is nagyon **szeretem**. Főleg a dél-európai filmek érdekelnek. '**I** also **like films a lot**. I am especially interested in films from Southern Europe.' **Szeretem** *megfigyelni az* állatokat és *a növényeket*. 'I like to observe animals and plants.'		

Table 6 – Occurrences with the verb *szeret* 'like, love'
related to various topics in the video corpus.

The thematically organised and corpus-based presentation of language forms is an important new way of developing study material. The thematic observation and analysis of language patterns reveals the meaning and use of lexical units, their grammatical preferences linked to a given topic. Each grammatical pattern occurs with a limited number of lexical items, and each lexical item follows a limited number of grammatical patterns (Hunston, Francis 2000, 83-146). And we have seen above that we can also organise grammatical patterns thematically. This does not imply exclusivity, but only that we are looking at the themes (and not at themes as illustrations of certain grammatical elements, or at sets of sentences that are not thematically connected at all). We are highlighting the linguistic elements that are often associated with a theme.

The task of unravelling the relationship between frequency of occurrence and topic – especially at beginner levels – is primarily the task of the language teacher and/or the author of the course material, and should be made available to students in a didactic form. The patterns that emerge may be different, and even material drawn from different corpora may differ. But they are always based on authentic linguistic manifestations and convey to the students the use of the language of production, which at the same time develops linguistic competence. And the 'discovery' and deductive processing of corpus-based and thematically organised collections can be an integral part of the classroom or of home preparation (the possibilities of which will be explained in another article).

The large linguistic input that appears thematically according to one linguistic element provides scope for observation, memorisation and provides a model for natural language use.

6. Summary and outlook

Visual input is part of most real-life interactions. It makes the situational context clear, shows which gestures accompany the utterances, and makes it clear what objects the speakers are referring to. The video recordings to *MagyarOK* are situated somewhere between semi-authentic and authentic data. The actors did not receive a script, only a brief outline of the situation that they were to improvise. They played several variations of the same interaction and were instructed to talk naturally but to avoid longer monologues.

Video corpora can form an integral part of both in-class learning and self-study by providing complex, natural-sounding, and relevant models at the learners' proficiency level. The students can observe these texts and adapt them to their needs, so that the texts can serve as models for their own text production. The *MagyarOK video corpus* is aimed at lower proficiency levels (A1-B1) and its key purpose is to add variations to the course book texts in order to increase the amount of linguistic input accessible to the learners. The recordings can also be used to illustrate the role of non-verbal communication on the one hand and the functions of selected lexico-grammatical elements on the other. This threefold focus on (1) the content of each text and (2) the linguistic and non-linguistic elements that make up

the content, and (3) the repetition and variation of texts, makes working with video recordings a meaningful activity in the classroom practice of language learning. The thematically structured teaching materials, fed by corpora, focus the language learning process on the association of topics with typical formal representations and on the preference of topics for certain grammatical elements.

However, further studies would be needed to measure the concrete, long-term benefits of using video corpus constructed in the way presented in this paper.

Works cited

Aston, Guy. 2001. *Learning with Corpora*. Houston: Athelstan.

Bennett, Shirley, Debra Marsh. «Are we expecting online tutors to run before they can walk?». *Innovations in Education and Teaching International* vol. 29, n. 1 (2002): 14-20. DOI: <https://doi.org/10.1080/13558000110097055>.

Braun, Sabine. 2006. «Getting Past 'Groundhog day': Spoken Multimedia Corpora for Student-centred Corpus Exploration». In *Corpus Linguistics in Language Teaching*, edited by Tony Harris and Maria Moleno Jaén, 75-98. Frankfurt am Main: Peter Lang.

Cohen, Jonathan D. 2017. «Cognitive Control – Core Constructs and Current Consideration». In *The Wiley Handbook of Cognitive Control*, edited by Tobias Egner, 3-28. New Jersey: Wiley-Blackwell.

Coniam, David. «The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study». *System* vol. 27, n. 1 (2001): 1-14. DOI: <https://doi.org/10.1016/S0346-251X(00)00057-9> (*open access*).

Egner, Tobias, Joy Hirsch. «Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information». *Natura Neuroscience* vol. 8 (2005): 1784-1790. DOI: <https://doi.org/10.1038/nn1594>.

Gabrielatos, Costas. « Corpora and Language Teaching: Just a fling or wedding bells?». *Teaching English as a Second of Foreign Language* vol. 8, n. 4 (2005): A-1. URL: <https://tesl-ej.org/ej32/a1.html> (*open access*).

Hunston, Susan, Gill Francis. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Krashen, Stephen D. 1988. *Second Language Acquisition and Second Language Learning*. New York: Prentice-Hall International.

Mehrabian, Albert. (1971) 1981. *Silent Messages: Implicit Communication of Emotions and Attitudes*. Belmont, CA: Wadsworth.

Oxford, Rebecca L., Madeline Ehrman. 1992. «Second Language Research on Individual Differences». *Annual Review of Applied Linguistics* vol. 13 (1992): 188-205. DOI: <https://doi.org/10.1017/S0267190500002464>.

Pelcz, Katalin. «Bepillantás a MID oktatásmódszertanának digitális átalakulásába: didaktizált videók a modellalapú oktatásban» [An insight into the digital transformation of HFL's educational methodology: didactic videos in model-based education]. *Hungarológiai Évkönyv* vol. 21, nn. 1-2 (2020): 123-131. URL: <https://epa.oszk.hu/02200/02287/00021/pdf/>.

Róka, Jolán. 2002. *Kommunikációtan*. [Communication studies]. Budapest: Századvég Kiadó.

Szirmai, Mónika. 2005. *Bevezetés a korpusznyelvészetbe. A korpusznyelvészet alkalmazása az anyanyelv és az idegen nyelv tanulásában és tanításában*. [Introduction to Corpus Linguistics. Application of corpus linguistics in learning and teaching mother tongue and foreign language]. Budapest: Tinta Könyvkiadó.

Szita, Szilvia. «Korpuszépítés és korpuszhasználat alacsonyabb nyelvtudási szinteken» [Corpus building and corpus use at lower language proficiency levels]. *Hungarológiai Évkönyv* vol. 21, nn. 1-2 (2020): 173-179. URL: <https://epa.oszk.hu/02200/02287/00021/pdf/>.

—. «A *MagyarOK* nyílt korpusz használatáról» [On the use of the *MagyarOK* open pedagogical corpus]. *Hungarológiai Évkönyv* vol. 22, nn. 1-2 (2021): 72-88. URL: <https://epa.oszk.hu/02200/02287/00022/pdf/>.

Szita, Szilvia, Katalin Pelcz. 2013-2019. *MagyarOK A1+, A2, B1, B2+ (Magyar nyelvkönyv + Nyelvtani munkafüzet* [*MagyarOk* Hungarian coursebook and workbook, CEFR level A1+, A2+, B1+, B2+]. Pécs: University of Pécs.

—. «Modellalapú nyelvoktatás. Természetes nyelvhasználat a tanteremben és a tantermen kívül» [Model-based language teaching. Natural language use in and out of the classroom]. *THL₂: Journal of Teaching Hungarian as a second language and culture* vol. 1, n. 2 (2017): 262-269. URL: <https://epa.oszk.hu/01400/01467/00015/pdf/>.

—. 2019-. «*MagyarOK* videókorpusz» [*MagyarOK* video corpus]. *MagyarOK - Magyarul beszélni menő!* [*MagyarOK* - Speaking Hungarian is cool!], YouTube channel, <https://www.youtube.com/channel/UCv4OI7fdvsdcXhe8DPyyQrg>.

—. 2020-. *«MagyarOK* Teaching Materials for Hungarian, Levels A1 to B2» (old version). URL: <https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fmagyarok> (*open access*).

—. 2022. *MagyarOK A1 Kompakt. Magyar Nyelvkönyv és Munkafüzet egy kötetben* [*MagyarOK* Compact. Hungarian coursebook and workbook pack, CEFR level A1]. Pécs: University of Pécs and Institute for Model-based Language Teaching.

White, Cynthia. 2003. *Language Learning in Distance Education*. Cambridge: Cambridge University Press.

Widdowson, Henry G. 2003. *Defining Issues in English Language Teaching*. Oxford: Oxford University Press.

Willis, Dave. 2003. *Rules, Patterns and Words. Grammar and Lexis in English Language Teaching*. Oxford: Oxford University Press.