

OSZKÓ BEATRIX – TÓTH NOÉMI EVELIN – YANG ZIJIAN GYŐZŐ

AZ ÁLTALÁNOS ÉS KÖZÉPISKOLAI MAGYAR NYELVTAN TANANYAG ELSAJÁTÍTÁSÁT SEGÍTŐ ALKALMAZÁS

1. Bevezetés

A 2020-ban kialakult vírushelyzet miatt Magyarországon is egyik napról a másikra kellett jelenléti oktatásról távoktatásra váltani országszerte (Szűts 2021). Emiatt sok család kényszerült arra, hogy részt vegyen gyermeke tanításában, ez sok esetben nem bizonyult egyszerű feladatnak. Ez a helyzet új kihívásokat állított tanárok, szülők és gyermekek elé egyaránt (Jánk–Lőrincz 2021). Segítségképpen egy olyan alkalmazás fejlesztését tűztük ki célul, mely az internet adta lehetőségeket használja az oktatásban, és használatával a magyar nyelvtan iskolában oktatott szabályai könnyebben elsajátíthatók. Reményeink szerint applikációnk használatával a diákok játékosan ismerkedhetnek meg azzal, hogy hogyan épülnek fel a mondatok, milyen egységekből állnak, és ezek hogyan viszonyulnak egymáshoz. A 2020-as átállás megmutatta azt is, hogy miért van szükség jól felépített, otthoni tanulás során is megfelelő informatikai eszközökre és alkalmazásokra, amelyekkel ki lehet egészíteni az általános és középiskolai oktatást, és amelyek ugyanakkor biztosíthatják a megfelelő kommunikációt tanár és diák között. Ezzel az alkalmazással szeretnénk csatlakozni a digitális oktatás fejlesztésére irányuló törekvésekhez, hogy a gyerekek a tanórán kívül is kaphassanak szakszerű segítséget, amivel biztos tudást építhetnek, és jól teljesíthetnek az órákon is. Mindezt játékos formában, hogy a tanulás öröm legyen, a diákok ne veszítsék el motivációjukat, és sikeresen teljesítsék a tantárgyi követelményeket. Mindezt tudományos igényességgel tettük, de első helyre a felhasználói igényeket helyeztük.

2. Elméleti megalapozás

2.1. Természetes nyelvek és mondatelemzés az iskolában

A kommunikációs eszközök rohamos fejlődése hatással van a mindennapi kommunikációkra is. Különböző szövegek vesznek minket körül az iskolában, a munkahelyen, az utcai hirdetőtáblákon és a világhálón is. A lista végtelen. Ezek segítségével tájékozódunk és nyújtunk tájékoztatást mások számára. A hangok, amiket hallunk, a szövegek, amiket olvasunk, mind olyan információt tartalmazó egységek, amelyeket képesek vagyunk megérteni és feldolgozni. A természetes nyelv megnyilatkozásai sokszor nem egyértelműek, szabályai nehezen írhatók le matematikai módszerekkel. A leírást/szabályalkotást tovább nehezíti, hogy a nyelvek változatokban léteznek. Hiába használjuk

tökéletesen anyanyelvünket, a nyelvi összefüggések formális leírása már nehéz feladat. Ezeket az összefüggéseket próbálja a nyelvten számunkra érthetően megmutatni.

Ahhoz, hogy modellünk az oktatásban is alkalmazható legyen, az ott használt szabályokat és alapfogalmakat kell ismerni. A nyelvtudomány eredményei és az anyanyelvi oktatás között nem egyszer jelentős különbségek vannak (Tolcsvai Nagy 2002). Ezekkel a különbségekkel ez a dolgozat nem foglalkozik, munkánk során a magyar közoktatásban tanított szabályokat alkalmazzuk, ezt hagyományos módszertannak nevezzük, és így hivatkozunk rá a későbbiek során.

Az iskolai nyelvten minden nyelvi szintet érint. A mondattant az alsóbb nyelvi egységek – a fonémák, morféimák, lexémák és szintagmák – után tárgyalja (Szabó 2010). A mondatot a szöveg alapegységének tekinti, erre épül a szövegtan és a pragmatika. Az iskolában oktatott mondattan függőségi szemléletű, a mondatot alkotó szavak, szintagmák egymáshoz való viszonyának megállapítása a célja. A tananyag szerint „a mondat egy vagy több szóból álló zárt intonációs szerkezet, a beszéd legkisebb egysége” (Fráter 2012: 96). A mondatok sokféleképpen osztályozhatók, például szerkezet vagy logikai minőség szerint. Dolgozatunkban az egyszerű mondatok felépítésével és elemzésével foglalkozunk. A mondatok szószerkezetekre bonthatók, ezeket két szó közötti meghatározott nyelvtani viszony hozza létre. A mondat fő részei az alany és az állítmány, ezek alkotják nyelvtani, jelentéstani és logikai magját. Egyeztetjük őket számban és személyben. E két fő egységen kívül az elemzés kiterjed további elemekre is, ezek a bővítmények: a tárgy, a határozók és a jelzők. E két utóbbi csoportnak a tankönyvekben előforduló típusai láthatók az 1-es és 2-es táblázatban, példamondattal kiegészítve.

Jelzők fajtája	Példa	Kérdőszó
minőség	piros alma	Milyen? Melyik? Hányadik?
mennyiség	tenyényi földdarab	Mennyi? Hányadrész?
birtokos	szomszédom kertje	Kinek a? Minek a?
kijelölő	ez a diák	Melyik?
értelmező	frisset, finomat	–

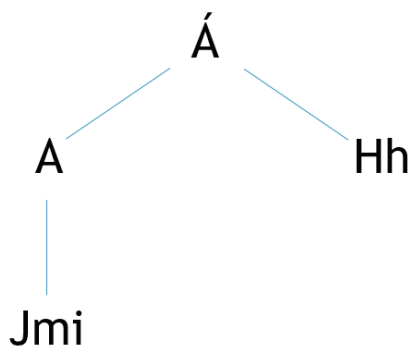
1. táblázat: A jelzők típusai (Fráter 2012)

Határozók fajtája	Példa	Kérdőszó
hely	a folyó mentén	Hol? Hová? Honnan? Merről?
idő	ma jött	Mettől? Mikor? Meddig? Mióta?
szám	gyakran látja	Hányszor?
állapot	álomba ringat	Milyen állapotban? Milyen állapotból/állapotba?
eredet	a sajt tejből készül	Miből?
eredmény	csonttá fagy	Mivé?
társ	a barátjával sétál	Kivel/Mivel együtt?
mód	angolul beszél	Milyen módon?
ok	gyötrődik valamiért	Mi okból? Mi miatt?
cél	küzd az ötösért	Mi célból? Mi végett? Mire? Mi célra?
tekintet	szellemileg friss	Minek a tekintetében? Milyen szempontból?
fok-mérték	átlagon felül teljesít	Mennyire? Milyen fokban/ mértékben?
eszköz	fuvolán játszik	Mi által? Milyen eszközzel?
részes	a diákok számára készült	Kinek/Minek a részére?
hasonlító	magasabb nála	Kinél? Minél?
állandó	büszke vagyok rá	–

2. táblázat: Határozótípusok (Fráter 2012)

Az alany, az állítmány, a tárgy, a határozók és a jelzők azok a mondatrészek, amelyeket egy mondat elemzése során fel kell ismernünk. Elemzéskor az iskolában az egyes mondatrészeket különbözőképpen jelölik, és a mondat szavai között megállapított viszonyokról sokszor ágrajzot is készítenek, melyek az alá-fölé rendeltségi viszonyokat ábrázolják. A mondatelemzésnél használt típusjelölő vonalakat rendszerint a típus kategóriájának kezdőbetűjével vagy jól érthető rövidítéssel jelölik. A mondatelemzés iskolában elfogadott módját az 1. ábra szemlélteti.

A_{mi} hatalmas jegesmedve az h_{északi-sarkon} él.



1. ábra: Egy elemzett mondat és ábrázolása

Alkalmazásunkban Fráter Adrienne 2012-ben megjelent *Magyar nyelv a középiskolások számára 9* című könyvét dolgoztuk fel, ezt vettük alapul a leckék összeállításához, ebből származnak példamondataink, az itt található nyelvtani szabályokat fogadtuk el. Az így létrejött törzsanyagot a következő tankönyvekből egészítettük ki: Antalné Szabó Ágnes és Raácz Judit *Magyar nyelv és kommunikáció 7.* (2012), Thomán Angéla *Magyar nyelv és kommunikáció. Tankönyv a 7. évfolyam számára* (2017) könyve. Felhasználtuk még a *Grammatikai gyakorlókönyv* (P. Lakatos, 2006) anyagát is.

2.2. A nyelvtechnológiáról

Már az 1950-es években megjelent az igény a nyelvek számítógéppel történő feldolgozására (Turing 1950/2009). Eleinte különböző nyelvű szövegek angolra fordítását szerették volna számítógépekkel elvégeztetni, azonban ez túl nagy kihívásnak bizonyult. Viszont ennek köszönhetően kezdődött el a számítógépes nyelvfeldolgozás (Nadkarni–Ohno-Machado–Chapman 2011), ami ma már minden olyan törekvést magába foglal, melynek célja, hogy a számítógép képes legyen az emberi nyelveket intelligensen feldolgozni és megérteni. A terület az utóbbi években indult rohamos fejlődésnek, ehhez az internet megjelenése, a rajta fellelhető megannyi szöveges adat, valamint a mesterséges intelligencia fejlődése is hozzájárult. Lényege, hogy csökkentsük a számítógép és az ember közötti kommunikációs szakadékot, azaz ne csak programnyelveken, hanem a természetes nyelvekhez minél közelebb álló nyelven lehessen a gépekkel kommunikálni. A nyelvtechnológiát sokféleképpen lehet használni, csak néhány ezek közül: gépi fordítás, szövegek különböző szempontok szerinti elemzése, helyesírás-ellenőrzés, de találkozhatunk vele a telefonunkba épített okosasszisztens használatakor, és ide tartozik a böngészők keresőmotorja is.

Amikor egy olyan számítógépes program készül, melynek célja az emberi nyelv megértése, különböző módszereket és eszközöket használhatunk a szövegfeldolgozáshoz, érdemes a célunknak leginkább megfelelő nyelvi szinten feldolgozni anyagunkat (Chowdhury 2003). Léteznek módszertanok, melyek kifejezetten egy adott nyelvi szint feldolgozását tűzik ki célul. Vannak lexikális, morfológiai és szemantikai elemzők. Azonban ezek használata előtt a szövegek előfeldolgozására van szükség ahhoz, hogy az információkat megfelelő módon nyerhessük ki. A szövegeket olyan formátumra kell hozni, mely alkalmas a számítógépes elemzésre. Ez leggyakrabban az annotáció készítése (Király 1970). Annotálás során a szöveget alkotó egységek általános jellemzőit adjuk meg. Az annotációs szerkezetekben a sorok egy-egy szót és annak a tulajdonságait írják le. Ezt ismeri fel és értelmezi a nyelvi elemző modell.

Bár a legnagyobb eredményeket általában angol nyelven érik el, számos magyar törekvés van arra, hogy a nyelvtechnológia eszközeit magyar nyelvű szövegeken is alkalmazni tudjuk. A nyelvi szintek feldolgozásához több eszköz közül választhatunk. Ilyen például a Quntoken (W1), mely mondatra bontó, azaz tokenizáló eszköz, és a Quex lexikális elemzőre épül, és C++ nyelven írták. A Humor névre elkeresztelt morfológiai elemző (Novák 2003) a magyar szóalaktani és szövegelemzéshez járul hozzá. Hasonló, bár más megközelítést alkalmaz a Magyarlánc (W2), egy szintaktikai elemző (Zsibrita–Vincze–Farkas 2013), mely a legalapvetőbb nyelvi szövegfeldolgozást végzi el magyar nyelvű korpuszokon. Talán a legismertebb magyar nyelvű alapon nyugvó nyelvfeldolgozó rendszer az e-magyar (W3), mely 2016-ban a magyar nyelvtechnológiai közösség összefogásával jött létre a Magyar Tudományos Akadémia támogatásával (Váradi et al. 2017). Ez egy integrált pipeline-t alkalmazó szoftver, azaz a rendszer képes egybegyűjteni, egységes láncba integrálni és közzétenni az elemzési lépéseket megvalósító magyar nyelvfeldolgozó eszközöket. Ennek eredményeképpen közvetlenül felhasználhatóvá válnak ezek az eszközök a különféle igényű felhasználói körök számára.

2.3. A gamifikáció

Az utóbbi években jelent meg az oktatás minden szintjén – és elsősorban a 10–18 éves korosztály körében igen sikeres – egy új motivációs módszer. A gamifikáció, más néven játékosítás a játékok és játékelemek alkalmazását jelenti az élet játékon kívüli területein (Deterding et al. 2011), célja pedig, hogy az ott zajló folyamatokat érdekesebbé és eredményesebbé tegye. A gyerekek játék közben fejlődnek és tanulnak a legtöbbet. A játékok eszközeit felhasználva utánozzák a való életet, ismerik fel a szabályokat és alkalmazkodnak hozzájuk. Ez a tendencia körülbelül 10 éves korukig tart, ezután a tanulás munkává, az iskola pedig munkahellyé válik a számukra. A játék csupán szabadidős tevékenységként marad meg az életükben, amit az emberek nagy része felnőttként is rendszeresen űz (Rigóczki 2016). A játékokban rejlő potenciál kihasználására találták ki a játékosítást, amely ma az élet számos területén elterjedt módszer a produktivitás növelésére

(Kovács–Várallyai 2018). Találkozhatunk vele a nagy cégektől kezdve az egészségügyön és az oktatáson át (Fromann–Damsa 2016) egészen a bevásárlóközpontok kedvezményeit népszerűsítő pontgyűjtő kuponokig.

A gamifikáció lényege, hogy megteremtse a felhasználó belső motivációját a továbbhaladásra és a hosszú távú teljesítésre. Mindezt különböző mechanizmusok és játékelemek segítségével éri el. A játékmechanizmusok alatt működési elveket kell érteni, a játékelemek pedig azok az eszközök, amelyek ezeket megvalósítják a gyakorlatban. Ezek hasonlítanak a játékok általános szabályrendszeréhez, ami nem is meglepő, hisz a játékok adta lehetőségek ihlették magát a gamifikációt is.

Mechanizmusok	Játékelemek
Önkéntesség	Narratíva
Elemi részekre bontott feladatok	Szintek
Állandó, azonnali visszacsatolás	Küldetések
Negatív stressz kiszűrése	Jelvények
Transzparencia	Pontozási rendszer
	Toplisták
	Látványelemek

3. táblázat: Mechanizmusok és játékelemek a gamifikációban

Ezeknek az elemeknek a megjelenése és arányuk egy-egy játékosítást alkalmazó rendszerben nagyban függ a célközönségtől és a játék tárgyától is. A jól gamifikált szoftver sikere pedig azon múlik, hogy mennyire jól adja vissza a játékok adta élményt.

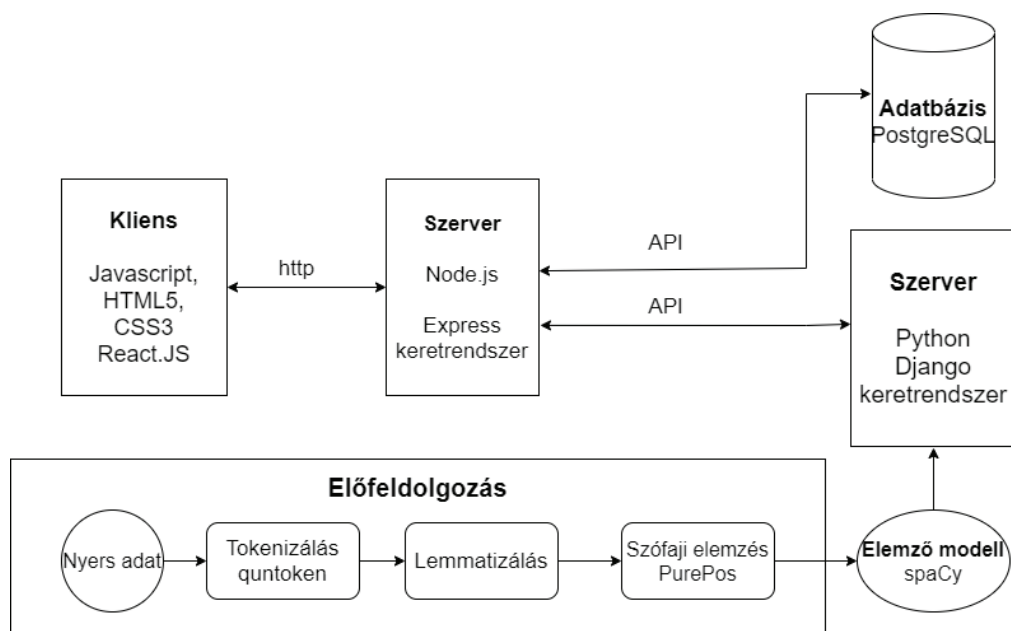
A gamifikációs elemek alkalmazása mára népszerűvé vált a mindenki számára elérhető szoftverekben is. Az oktatóalkalmazások területén mind közül a legismertebb a Duolingo (W4) idegen nyelvet tanító applikáció. Több mint 50 nyelv elsajátítására kínál lehetőséget, könnyen használható, és sok gamifikációs elemmel rendelkezik. A tanulás során lingotok gyűjthetők, az eredmények jelvények formájában jelennek meg. A tanulás személyre szabható, a felhasználó célokat állíthat fel magának, meghatározhatja, hogy naponta mennyi időt szeretne tanulással tölteni, és az alkalmazás segít ezt betartani. A Duolingo 2019-ben 30 millió (W5) aktív felhasználóval rendelkezett. Ezenkívül sok e-learning oldal és tanulóalkalmazás szintén gamifikációs alapokon nyugszik, mint például a Udemy (W6) és a Khanacademy (W7), melyek online kurzusokat kínálnak szinte bármilyen témában. A Remind (W8) iskolás osztályoknak létrehozott közösségi platform, a Kahoot (W9) egy nagyon népszerű kvízzjátékkészítő alkalmazás. Az ezzel összeállított kvízeket a diákok az órán kitölthetik, majd a program rangsorol a teljesítmény alapján. A mi alkalmazásunk leginkább a magyar fejlesztésű Matekingre (W10) hasonlít. A 2011-ben publikált, középiskolai és egyetemi matematikát oktató e-learning

oldal (Viktor–Kárpáti–Daróczy 2020) színes és igényes, rendkívül aprólékos ábrákkal magyarázza a matematikai szabályokat, így könnyen érthetővé és befogadhatóvá válik az információ.

3. A gyakorlati megvalósítás

3.1. Az alkalmazás

A rendelkezésünkre álló elméleti anyag áttekintése után kezdődött meg az alkalmazás fejlesztése a legkorszerűbb, nyílt forráskódú technológiákat felhasználva. Az alapok a front-end Javascript (W11) nyelven készültek, ezt használja a weboldalak 94,5%-a, és a legtöbb böngésző támogatja. Mindezt a Javascript React (W12) könyvtárával egészítettük ki, hogy a felhasználói felület még jobb minőségű legyen. Célunk olyan felület biztosítása, mely használatát a kisebb gyermekek és a szülők is könnyedén és gyorsan megtanulják, és ami kényelmesen elérhető, használható bármilyen eszközről, legyen szó számítógépről, mobiltelefonról vagy tabletről. Az adatok alsóbb rétegen való feldolgozását a Node.js Express (W13) keretrendszere biztosítja, mellyel Javascript kódot backenden írhatunk. Ez a rendszer az általunk megírt API-n keresztül kommunikálja az adatokat a front-end és a hozzácsatolt PostgreSQL (W14) adatbázis között. Erre az adatbázisra azért van szükség, hogy a felhasználók adatait egyszerűen és biztonságosan tárolhassuk. Az alkalmazás architektúráis terve a 2. ábrán tekinthető át.



2. ábra: Az alkalmazás architektúráis terve

A weboldalon a diákok saját felhasználót regisztrálhatnak, ehhez készül egy felhasználói profil, ahol az eredményeik – a leckék teljesítésével összegyűjtött pontok és a különböző kihívások teljesítéséért kapott jelvények – található meg. Ezenkívül itt láthatják és szerkeszthetik adataikat, melyek módosításkor azonnal frissülnek az adatbázisban. A felhasználóknak lehetőségük van a fiókjuk törlésére és jelszavuk cseréjére is. A rendszerben különböző jogosultsági körökkel rendelkezhetnek a felhasználók: lehet valaki adminisztrátor, tanár vagy diák. Az adminisztrátor kezeli az adatbázisban megtalálható adatokat, és ő az, aki a tanár jogosultságú felhasználókat regisztrálni tudja. Tanár jogosultsággal saját tesztfeladatok készíthetők, melyek megoszthatók a diákokkal. Regisztrációkor automatikusan diáknak sorol be mindenkit a rendszer. A diákok látják a tananyagokat és a hozzájuk tartozó feladatsorokat.

A témakörök követik az iskolai anyag felépítését, és példamondatokon keresztül mutatják be a mondatelemzést. A témakörök között egy térképen navigálhat a felhasználó. Hasonlóképpen működnek a tesztek is, melyek minden fejezetben vannak, és egy-egy 5-10 kérdésből álló feladatlapnak felelnek meg. A jó válaszokért cserébe pontokat ad a rendszer, ezek megsokszorozhatók, ha egymás után többször jó választ adunk, vagy ha az egész feladatsort hibátlanul sikerül teljesíteni. A megoldott tesztfeladatokat zölddel jelöljük. A feladatsorok szintekhez vannak rendelve, ha egy szinten minden feladatot megoldottunk, új szintet nyithatunk meg. A tesztek változatosságát különböző tesztípusok biztosítják, például kérdés-válasz, húzd a megfelelő helyre, kvíz, párosító, kiegészítendő szöveg. Ezek a felhasználói interakciót növelik. A tesztek tetszés szerint többször ismételhetők, a korábban megoldottak is. A rendszer külön figyeli a tanulók teljesítményét, és jelvényeket oszt ki, ha megoldottak például 10, 50 vagy 100 tesztfeladatot. A megoldások számát feltüntető jelvény az adatlapon is szerepel. A jelvények rendkívül változatosak lehetnek, és a szoftver lehetővé teszi a későbbi bővítést is.

3.2. Annotációkészítés, az elemzőmodellek tanítása

A számítógépes szövegfeldolgozásban a mondat szerkezetek reprezentálására az igeközpontú függőségi elemzés terjedt el. Ez a hagyományos leíró magyar nyelvtanhoz közel áll, ezért is választottuk a későbbi elemzések elkészítéséhez. Mind a függőségi, mind pedig az iskolai elemzés a mondatokat fa szerkezetben ábrázolja, kiindulópontjuk az állítmány. Lényeges különbség azonban, hogy míg a függőségi elemzés tokenekkel dolgozik, az iskolai nyelvtan szintaktikai szavakat használ, és az alany az állítmánynak bővítménye. Egy szintaktikai szó esetenként több tokenből is állhat, a hagyományos elemzés azonban nem állapít meg további viszonyt az egyes elemek között. A fában ezek egyszerűen több szóból álló csomópontként jelennek meg. Ilyenek például az összetett igealakok (*lett volna, menni fog*), a többtagú nevek, vagy az igekötős igék. A függőségi elemzésben a mondat elemei külön tokeneket alkotnak, a névelők és

az írásjelek is, azonban az iskolai elemzés során ezekkel és a többi funkciószóval nem foglalkozunk. Ezenkívül a hagyományos elemzési fában nem különül el egymástól a predikatív névszó és a kopula sem, ezek együtt alkotják az úgynevezett névszói-igei állítmányt. Ezeket a különbségeket mindenképpen figyelembe kellett venni a címke-készlet megvalósítása során.

Az elemzőmodell elkészítéséhez a spaCy elnevezésű szövegfeldolgozó eszközt választottuk. Ez ingyenes, és könnyen megtanulható a használata. A spaCy konvolúciós neurális hálókat használ a szövegfeldolgozáshoz, ami azért is jó, mert így egyszerre több adatot „lát”, és párhuzamosan tudja feldolgozni a kapott bemeneti tokeneket. Az új nyelvi modellek is – mint például a Bert – mind ezzel a módszerrel dolgoznak. Jelenleg a spaCy az egyik leggyorsabb szintaktikai elemző, és csak közel 1%-kal marad le a legjobban teljesítő rendszerektől. Python nyelven írták, és a természetes nyelvfeldolgozás különböző területein nyújt segítséget. A hasonló programokhoz képest inkább ipari, praktikus megközelítésű. Ez előnyös tulajdonság, ha azt szeretnénk, hogy termékünk a lehető legrövidebb idő alatt eljusson a felhasználókhoz. A spaCy-vel tehát egyszerűen végezhetjük el például szövegek előfeldolgozását, tokenizálását, szavak szótövének kinyerését, szövegek osztályozását, szófajcímkézését. Mi ez utóbbit használtuk. Külön érdekesség, hogy a spaCy pipeline-ja – mely a szövegfeldolgozást végzi – igényeinkre szabható, új komponensekkel bővíthetjük, és dönthetünk úgy, hogy egyes komponenseket nem használunk fel.

A spaCy-nek még nincs előre letölthető saját, magyar nyelvű nyelvi csomagja, ezért Orosz György magyar modelljével (W15) dolgoztunk. A modell a Szeged Treebanken (W16) alapul, ez ma a legnagyobb teljesen kézzel annotált magyar nyelvű adatbázis. Következő lépésként az általunk készített címkekészlettel kezdtük meg a modell tanítását, mely a tanítóanyag szavaiból szövektorokat képezve hozott létre egy neurális hálót, ami képes volt a példák segítségével a tesztmondatoknál eldönteni, milyen mondatrész az adott szó, és mely szóval áll függőségben. Az összegyűjtött és kézzel annotált mondatokat .conll fájlformátumban adtuk át. A tanítóanyag 268, a tesztanyag pedig 50 kézzel elemzett mondatból állt. A spaCy pipeline komponenseiből a parsert és a taggert alkalmaztuk. Ezek elegendőnek bizonyultak ahhoz, hogy a modell a szavakhoz címkét tudjon rendelni.

3.3. Az eredmények kiértékelése

A pontosabb eredmények érdekében két modellt is elkészítettünk, melyek között egyetlen eltérést a címkekészlet részletessége jelentette. A kis modell esetében csak a legalapvetőbb címkéket használtuk fel, mint az állítmány, alany, tárgy – és nem részleteztük a különböző határozó- és jelzőtípusokat. A nagy modellnél ez utóbbi két kategória altípusait is megadtuk. A modellek eredményei a 3. és 4. táblázatban láthatóak.

Mondatrész	Pontosság	Fedés	F–mérték	Gyakoriság a tesztadatban
Állítmány	96.00%	92.31%	94.12%	52
Alany	70.83%	77.27%	73.91%	22
Tárgy	90.00%	83.72%	86.75%	40
Határozó	75.68%	82.35%	78.87%	34
Jelző	93.75%	65.22%	76.92%	23
P	51.85%	66.67%	58.33%	21
X	95.96%	95.96%	95.96%	99

3. táblázat: A kis modell eredménytáblázata

Mondatrész	Pontosság	Fedés	F–mérték	Gyakoriság a tesztadatban
Állítmány	92.00%	88.46%	90.20%	52
Alany	76.19%	72.72%	74.42%	22
Tárgy	76.32%	72.50%	74.36%	40
Minőségjelző	71.43%	62.50%	66.67%	8
P	55.17%	76.19%	64.00%	21
X	97.98%	97.98%	97.98%	99

4. táblázat: A nagy modell eredménytáblázata

Az eredmények alapján beláttuk, hogy a nagyobb modell ritkán előforduló címkéinek sikeressége, azaz a pontosság és a fedés nem kiértékelhető, ezért ezeket nem is mutatjuk a 4. táblázatban. Ebben a modellben az alapcímkéken kívül további 16 címke fordult elő (többnyire különböző határozók) a tesztmondatokban, az eredményben azonban ezek egyszer sem szerepeltek. Ez azt mutatja, hogy a tanítókorpusz méretéhez képest túl nagy volt a címkékészlet, emiatt túl sok volt a ritka címke. Ezért tettünk kísérletet a szűkített címkékészletű kis modellel is, amely nem foglalkozott a különböző határozó- és jelzőtípusokkal. Nem meglepő módon ez az egyszerűbb modell eredményesebbnek bizonyult. A címkékészlet egyszerűsítése kapcsán azonban ki kell emelni a P címke feltűnően alacsony eredményeit. Ez a címke minden olyan szót magában foglalt, amely valamelyik másik szóval együtt alkot egy mondatrészt. Tartoztak ide igekötők, vezetéknevek, segédigék és predikatív névszók is. Úgy tűnik, nem volt szerencsés ezeknek az összevonása, és a későbbiekben érdemesebb lesz az (egyébként rendelkezésre álló) specifikusabb címkét használni. Ezzel szemben a nem elemzendő tokenek (névelők, írásjelek, partikulák stb.) egyetlen X címkébe vonása sikeresnek bizonyult. A függőségek értékelésénél – mint korábban említettük – nem vettük figyelembe az X-szel jelölt, nem elemzendő szavakat. Mindkét modell nagyjából 81%-os pontossággal jósolta meg az elemzendő szavak fejét, azaz a függőségeket.

4. Továbblépési lehetőségek

A kutatás tapasztalatai alapján elmondható, hogy ha tovább bővítjük a tanítóanyagunkat, és jobb lefedést biztosítunk a példamondatokkal a különböző jelzők és határozók használatára, az eredményeink tovább javíthatók, és ez növeli annak a valószínűségét, hogy az elemzőmodell biztonsággal alkalmazható lesz oktatási célokra. Mindez a módszertan további finomítását követeli meg.

5. Összefoglalás

Dolgozatunkban egy, az iskolai mondatelemzést segítő applikáció fejlesztésének első lépéseit mutattuk be. Mivel a mai, a közoktatásban részt vevő fiatal generációnak nem jelent különösebb gondot az internetes eszközök használata, alkalmazásunkat érdemes lesz integrálni a tanítási folyamatba egyrészt azért, hogy az önálló tanulást megkönnyítsük, másrészt azért, hogy a pedagógusok a tanórákat színesebbé tehesék. Reméljük, hogy ezzel a szoftverrel hozzájárulunk majd ahhoz, hogy a magyar nyelvtant megszeressék az általános és középiskolás diákok.

Irodalom

- Chowdhury, Gobinda 2003. Natural language processing. *Annual review of information science and technology* 37/1: 51–89.
<https://doi.org/10.1002/aris.1440370103>
- Deterding, Sebastian – Dixon, Dan – Khaled, Rilla – Nacke, Lennart 2011. From game design elements to gamefulness: defining „gamification”. In: *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. New York: Association for Computing Machinery. 9–15.
<https://doi.org/10.1145/2181037.2181040>
- Fromann Richárd – Damsa Andrei 2016. A gamifikáció (játékosítás) motivációs eszköztára az oktatásban. *Új Pedagógiai Szemle* 66: 76–81.
- Jánk István – Lőrincz Marina 2021. Hogyan zajlott a magyar nyelv digitális tanítása a járványhelyzet alatt? *Anyanyelv-pedagógia* 14/4.
<https://doi.org/10.21030/anyp.2021.4.1>
- Király István (főszerk.) 1970. *Világirodalmi lexikon I. (A–Cal)*. Budapest: Akadémiai Kiadó. 340.
- Kovács Tamás – Várallyai László 2018. Gamifikáció, avagy a játékosítás szerepe napjainkban. *International Journal of Engineering and Management Sciences* 3/3: 171–180.
<https://doi.org/10.21791/IJEMS.2018.3.14>
- Nadkarni, Prakash – Ohno-Machado, Lucinda – Chapman, Wendy 2011. Natural language processing: an introduction. *Journal of the American Medical Informatics Association* 18/5: 544–551.

- <https://doi.org/10.1136/amiajnl-2011-000464>
- Novák Attila 2003. Milyen a jó Humor. In: Alexin Zoltán; Csendes Dóra (szerk.): *Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem. 138–145.
- Rigóczki Csaba 2016. Gamifikáció (játékosítás) és pedagógia. *Új Pedagógia Szemle* 66: 69–75.
- Szabó Veronika 2010. *A magyar mondat modelljei a nyelvtanoktatásban*. Kézirat. Pécs: Pécsi Tudományegyetem Nyelvtudományi Doktori Iskola.
- Szűts Zoltán 2021. (Táv)oktatás a koronavírus idején. *Korunk* 3/2: 9–16.
- Tolcsvai Nagy Gábor 2002. A magyar nyelv leírása és iskolai oktatása. https://mta.hu/data/dokumentumok/i_osztaly/1_Eloadasok_tara/Magyar_nyelv_es_kutata-sa_20020502/TolcsvaiN_leiras_oktatas_20020502.pdf
- Turing, Alan Mathison 1950/2009. Computing machinery and intelligence. In: Robert Epstein, Robert – Roberts, Gary – Beber, Grace (eds.): *Parsing the Turing test*. Dordrecht: Springer. 23–65.
- Váradai Tamás – Simon Eszter – Sass Bálint – Geröcs Mátyás – Mittelholcz Iván – Novák Attila – Indig Balázs – Prószéky Gábor – Farkas Richárd – Vincze Veronika 2017. Az e-magyar digitális nyelvfeldolgozó rendszer. In: *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017)*. Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport. 49–60.
- Viktor Patrik – Kárpáti-Daróczy Judit 2020. Innovatív e-learning rendszerek elemzése. In: *Vállalkozásfejlesztés a XXI. században*. Kézirat. Budapest: Óbudai Egyetem Keleti Károly Gazdasági Kar.
- Zsibrita, János – Vincze, Veronika – Farkas, Richárd 2013. magyarlanc: A tool for morphological and dependency parsing of Hungarian. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing*. Hissar: RANLP. 763–771.

Források

- Antalné Szabó Ágnes – Raátz Judit 2012. *Magyar nyelv és kommunikáció* 7. Budapest: Nemzeti Tankönyvkiadó.
- Fráter Adrienne 2012. *Magyar nyelv a középiskolások számára* 9. Szeged: Mozaik Kiadó.
- P. Lakatos Ilona (szerk.) 2006. *Grammatikai gyakorlókönyv*. Budapest: Bölcsész Konzorcium.
- Thomán Angéla 2017. *Magyar nyelv* 7. Tankönyv. Eger: Eszterházy Károly Egyetem, Oktatókutatató és Fejlesztő Intézet.

Internetes hivatkozások

W1: <https://github.com/dlt-rilmta/quntoken> (2022. 02. 25.)

W2: <https://rgai.inf.u-szeged.hu/magyarlanc> (2022. 02. 25.)

W3: <https://e-magyar.hu/hu/> (2022. 02. 25.)

W4: <https://www.duolingo.com/> (2022. 02. 25.)

W5: <https://www.businessofapps.com/data/duolingo-statistics/> (2022. 02. 25.)

W6: <https://www.udemy.com/> (2022. 02. 25.)

W7: <https://www.khanacademy.org/> (2022. 02. 25.)

W8: <https://www.remind.com/> (2022. 02. 25.)

W9: <https://kahoot.it/> (2022. 02. 25.)

W10: <https://www.mateking.hu/> (2022. 02. 25.)

W11: <https://www.javascript.com/> (2022. 02. 25.)

W12: <https://reactjs.org/tutorial/tutorial.html> (2022. 02. 25.)

W13: <https://expressjs.com/> (2022. 02. 25.)

W14: <https://www.postgresql.org/> (2022. 02. 25.)

W15: <https://github.com/oroszy/spacy-hungarian-models> (2022. 02. 25.)

W16: <https://rgai.inf.u-szeged.hu/node/113> (2022. 02. 25.)