

MODRIÁN-HORVÁTH BERNADETT – KALIVODA ÁGNES –  
TANOS BÁLINT

**A magmondattal szerepű, passzív jelentésű sémák  
produktivitásának kvantitatív vizsgálata<sup>1</sup>**

**Kivonat**

A tanulmány célja négy passzív jelentésű konstrukció: a *V-ódi*k/-*ődi*k, a *V-(t)at*ik/-*(t)eti*k, a *V-va/-ve les*z és a *V-va/-ve van* konstrukció produktivitásának korpuszalapú vizsgálata. A kutatás az MNSZ2, valamint az újraannotált MNSZ2-UD korpuszokon végzett gépi és kézi elemzéseken alapul, és a kvantitatív keretben meghatározható globális produktivitási indexet (vö. Baayen 1992) kvalitatív elemekkel ötvözi. A globális produktivitási index mérése egy, a használatalapú nyelvészetben gyakran alkalmazott módszer, mely a konstrukciós sémák típus- és példánygyakoriságának, valamint a sémákkal képzett hapax legomenonok számának segítségével számítható ki. A vizsgálat kvalitatív része az egyes sémák produktív jelentését leginkább tükröző hapaxokból nyert minta igenemek szerinti besorolására irányult.

A vizsgálat eredményei azt mutatják, hogy a korpusz egészében a legnagyobb globális produktivitással rendelkező séma a *V-va/-ve van* konstrukció, a legnagyobb típusgyakorisággal a *V-ódi*k/-*ődi*k, míg a legnagyobb

---

<sup>1</sup> A dolgozat elkészítését Modrián-Horváth Bernadett részéről az NKFIH 129040 számú pályázata (A magyar nyelv igei konstrukciói. Használatalapú konstrukciós nyelvtani kutatás) támogatta. Kalivoda Ágnes a 4. fejezetben ismertetett korpuszvizsgálatot és a 6.2. fejezethez a random mintavételezést végezte. A kézi korpuszvizsgálatban a szöveg tisztítási feladatokat és a hapaxok arányainak számításait Tanos Bálint végezte. Modrián-Horváth Bernadett az 5. és 6. fejezetben bemutatott kézi korpuszvizsgálatokat kivitelezte, valamint a dolgozat elméleti részeit írta. Ezúton fejezzük ki hálánkat lektorainknak, Kugler Nórának és Palágyi Lászlónak, akik számos segítő és inspiráló megjegyzéssel, korrektúrával, részletes kommentárral segítettek jobbá tenni a tanulmányt. A fennmaradó hibák természetesen a szerzőket terhelik.

potenciális produktivitással a *V-va/-ve lesz* rendelkezik; az egyes alkorpuszokra vonatkozó eredmények ettől részben eltérnek. A hapaxokból nyert véletlenszerű minta elemzése alapján a legstabilabban passzív jelentésű séma a rendkívül alacsony produktivitású *V-tatik/-tetik* konstrukció, ezt szorosan követi a *V-va/-ve lesz*, a túlnyomórészt passzív jelentésű a *V-va/-ve van*, és a hapaxok kevesebb mint felében találunk passzív jelentést a *V-ódik/-ődik* esetében. A tanulmány további, elméleti és metodológiai megfontolásokat is tartalmaz.

**Kulcsszavak:** magyar nyelvtan, igenem, passzívum, korpusznyelvészet, konstrukciós nyelvtan

## 1. Bevezetés

### 1.1. Célkitűzés és elméleti háttér

A tanulmány négy, nem modális, részben vagy elsősorban passzív jelentésű,<sup>2</sup> magmondattal (állítmányi)<sup>3</sup> szerepű konstrukció kvantitatív produktivitási vizsgálatát tűzi ki célul. Konstrukció alatt általánosságban tanult, illetve itt különösen grammatikai típusú, azaz részben kitöltött vagy nem-kitöltött forma-jelentés párokat értünk (vö. Goldberg 2006, 2013). A vizsgált négy konstrukció: a *V-ódik/-ődik*, a *V-(t)atik/-(t)etik*, a *V-va/-ve lesz* és a *V-va/-ve van* konstrukció. A vizsgálatok elméleti alapvetése a használatalapú konstrukciós nyelvészethez kötődik (vö. Bybee 2010, 2013; Diessel 2014; Langacker 1987). A konstrukciós nyelvészet számára a bipoláris (fonológiai és szemantikai pólussal rendelkező) nyelvi jel fogalma nemcsak a morféma és lexéma alkalmazható, hanem különböző komplexitású és absztrakciós fokú konstrukciókat foglal magába (vö. Goldberg 2006, 2013), így a grammatika és a lexikon szerveződése kontinuumként modellálható. A használatalapúság azt tételezi fel, hogy az egyes konstrukciók a nyelvhasználati események során keletkeznek, a reprezentációk stabilizálódnak vagy éppen meggyengülnek. A példányalapú reprezentációk között felismert fonológiai és szemantikai

<sup>2</sup> A továbbiakban az egyszerűség kedvéért „passzív jelentésű konstrukciók” néven említjük a négy vizsgált szerkezetet, tekintet nélkül arra, hogy az aktuális konstrukciós jelentések milyen arányban képesek passzív jelentést kifejezni.

<sup>3</sup> A magmondattal fogalmához lásd Imrényi (2017).

kapcsolatok alapján a konstrukciók viszonya hálózatos. Az asszociatív viszony létesülése nem tudatos szinten zajlik, a konstrukciókon gyakorisági hatások érvényesülnek.

Amikor azt állítjuk, hogy egy nyelvben „van passzív”, ez alatt rendszerint azt értjük, hogy létezik az adott nyelvben egy vagy több olyan produktív konstrukciós séma, melynek segítségével cselekvő igei bázissal ágensdefokuszált perspektívából konstruálhatunk meg eseményeket, ezért a passzívumot mindig jelölt igenemnek szokták tekinteni. A passzív fogalmának meghatározására a jelen tanulmány céljainak megfelelően egy inkább operacionális definíció szolgál: passzívnak tekintjük a morfoszintaktikailag jelölt fókuszváltást egy alapige két fokális résztvevője között, más szóval azt az esetet, amikor az alapige jelentése szerinti elsődleges és másodlagos figura (trajektor és landmark) egy konstrukció nézőpontszerkezetében „felcserélődik” (vö. Langacker 2006).<sup>4</sup> A konstrukciós nyelvészeti megközelítés alapján a passzív kifejezőeszközök (is) alaki és szemantikai megfeleléseken alapuló, kapcsolódásokban gazdag hálózatot alkotnak (a hálózat modellálásához vö. Modrián-Horváth 2024). A jelen vizsgálatok pusztán metodológiai célból emelnek ki e hálózatból négy konstrukciót, miközben működésükre, elmebeli tárolódásukra egyaránt kihatással vannak a hálózat többi tagjai, így például a *V-ódik/-ődik* sémában is működő passzív jelentésű *-ó/-/ő-* (vö. Kugler-Simon 2018), a passzív jelentésű, kopula nélküli igenévi magmondatok (például *leves tálalva, küldetés teljesítve, a ház eladó*), a modális passzív konstrukciók (például *ehető*) vagy a nem magmondat szerepű passzív konstrukciók (például *Pontosz legyőzése*).

A passzív diatézis vizsgált kifejezőeszközeinek az inflexiós vagy szóképzési elemek közé való besorolása a konstrukciós nyelvészeti háttérnek megfelelően nem tartozik a jelen tanulmány fő kérdésfeltevései közé, legfeljebb annyiban, hogy az egyes konstrukciók produktivitási mértékét összehasonlítjuk egy tipikusan inflexiós jellegű analitikus konstrukció, a *V-ni fog* mutatóival.<sup>5</sup> Az inflexiós elemekre a szakirodalom szerint magas, míg a derivációs

<sup>4</sup> Langacker (2006) a passzív konstrukciót a „trajector-landmark reversal” egyik eseteként tartja számon.

<sup>5</sup> Ladányi (2017: 567kk., különösen 102. lj.) a *V-ni fog* sémára illeszkedő, „jövő idejű, összetett, segédigés alakok”-at (lásd Ladányi 2017: 567-568) morfológiai szavaknak tekinti, leírásukat a morfológia feladatának tartja a szóalakokkal való funkcionális hasonlóság (lehorgonyzás), valamint az erős grammatikalizáltság miatt. Palágyi László lektorunk e besorolást többek között a korábbi szakirodalom alapján sem tartja problémamentesnek (Palágyi 2024, jelen kötetben).

elemekre alacsonyabb produktivitás jellemző, így ez az összehasonlítás közvetett módon segíthet árnyalni a vizsgált konstrukciók hovatartozásáról alkotott képet (az analitikus szerkezetek esetében is).

A produktivitás több dimenzióban meghatározható, összetett fogalom (vö. Ladányi 2008; Baayen 2009). A jelen tanulmány – számos produktivitási kutatáshoz hasonlóan (vö. Baayen–Lieber 1991; Baayen 1992; Dressler–Ladányi 2000) – Schultink (1961) van Marle (1985) által angolra fordított definícióját veszi alapul, mely az eredeti változatban csak a morfológiai produktivitásra terjed ki: produktivitás alatt a nyelvhasználók számára rendelkezésre álló, alak–jelentés párok összefüggésein alapuló, elméletileg korlátlan mennyiségű képzési produktum szándékolatlan létrehozásának lehetőségét értjük.<sup>6</sup> A produktivitás méréséhez a kvantitatív szempontú globális produktivitást használjuk (lásd Baayen–Lieber 1991; Baayen 1992;<sup>7</sup> Zeldes 2012; bővebb kifejtését lásd a 2. fejezetben). A produktivitás ebben a felfogásban graduális jellegű, viszonyfogalom. A tanulmány kérdésfeltevése elsősorban a négy vizsgált konstrukció produktivitásának egymáshoz képest való elhelyezésére irányul.

## 1.2. Felépítés

Az 1–2. fejezet bemutatja a tanulmány kérdésfeltevést, az elméleti alapvetést és a használt korpusznyelvészeti módszert. A 3. fejezetben a passzív konstrukciók szemantikai tulajdonságait tárgyaljuk, ennek során kitérünk a konstrukciók polyszemiájára és a korpuszvizsgálat során jelentőséggel bíró elhatárolási nehézségekre is. A 4–7. fejezet az empirikus kutatásokról számol be. A 4. fejezetben elsőként bemutatjuk a vizsgálatához használt korpuszt (MNSZ2-UD) és az adatgyűjtés folyamatát. Ezután ismertetjük az eredményeinket, előbb a teljes korpuszra, majd az egyes alkorpuszokra vonatkozóan. Az 5. fejezet az MNSZ2 egy alkorpuszából, a beszélt nyelvi alkorpuszból származó adatok kézi elemzésének folyamatát írja le és annak eredményeit mutatja be. A 6. fejezet arra a kérdésre keresi a választ, hogy az elemzett konstrukciók mely sémajelentéssel fordulnak elő produktívan,

<sup>6</sup> „By productivity as a morphological phenomenon we understand the possibility for language users to coin, unintentionally, a number of formations which are in principle uncountable, by means of the morphological process which underlies the form-meaning correspondence of some words already known to them” (van Marle 1985: 100–101).

<sup>7</sup> A globális produktivitás meghatározása Baayen (1992)-ben két mutatóra is vonatkozik, az itt használt formula a  $P^* (n1/N; V)$ .

ehhez a hapaxokból nyert minta kvalitatív vizsgálatát, valamint a magas gyakoriságú elemek áttekintését végeztük el.

Az empirikus eredményekre a 7. fejezetben reflektálunk: összevetjük a produktivitási mutatókat a szakirodalomban megfogalmazottakkal, összehasonlítjuk a gépi és kézi vizsgálat eredményeit, következtetéseket vonunk le a hapaxok vizsgálatának eredményeiből és állást foglalunk a kutatás metodológiai újdonságával, a szintetikus és analitikus konstrukciók produktivitásának együttes vizsgálatával kapcsolatban. A tanulmányt összefoglalással és a nyitott kérdések megfogalmazásával zárjuk (8. fejezet).

## 2. A vizsgálati módszerről

A vizsgálat kvantitatív jellege kapcsolódik egyrészt a kötet célkitűzéséhez, azaz a magyar igei konstrukciók nyelvtechnológiai módszerekkel történő kutatásához, másrészt illeszkedik ahhoz a tágabb, kognitív nyelvészeti keretben megfigyelhető trendhez, mely szerint a nyelvészek a használatalapúságot szem előtt tartva egyre erőteljesebben támaszkodnak empirikus, kvantitatív vizsgálatokra (vö. Janda 2013).

A főként Baayen által kidolgozott kvantitatív produktivitási modelleket, illetve mutatókat (vö. Baayen–Lieber 1991, Baayen 1992, 1993) számos használatalapú, illetve konstrukciós nyelvészeti tanulmány alkalmazza (például Stefanowitsch 2006 egy esettanulmányban, Zeldes 2012 és Hilpert 2013). Ennek egyik oka az, hogy a használatalapú nyelvtanok egyik centrális feltevése szerint a gyakoriság (típus- és példánygyakoriság) fontos szerepet játszik az egyes konstrukciós sémák mint mentális reprezentációk létrejöttében, stabilizálásában és működési mechanizmusában. Ezenkívül a produktivitást mindkét megközelítési mód fokozati fogalomnak ismeri el, és nem diszjunkt kategóriaként kezeli (produktív vs. nem produktív vagy szemiproduktív), felismerve, hogy a produktivitás háttérben nagyfokú individuális és szociokulturális nyelvi komplexitás áll (vö. Baayen 2009). További közös vonás, hogy a kvantitatív modellálás és a használatalapú nyelvtan kiindulópontja egyaránt nem egy „ideális” beszélő kompetenciája, hanem a nyelvi adatokon alapuló nyelvreírás, mely a kvantitatív modellálás esetében mérhető is.

Baayen (1992) a kvantitatív modellálási lehetőségek bemutatását leginkább pszicholingvisztikai összefüggésekbe helyezi és a lexikális reprezentáció erőssége, az elérés, aktiváció felől magyarázza, kitér azonban Bybee (1985) használatalapú elméletére, illetve az abban feltételezett gyakorisági hatásokra

is. Alapjában véve Baayen egyetért a gyakoriságnak a használatalapú morfológiai modellálásban tulajdonított szerepével, de túlságosan leegyszerűsítőnek tartja a gyakoriság és a szabályosság kapcsolatát Bybee modelljében. Míg Bybee abból indul ki, hogy az alacsony és közepes gyakoriságú típusok szabályos képzési módokat reprezentálnak, és a magas gyakoriságúak szabálytalanok, Baayen (1992: 140) megjegyzi, hogy a magas gyakoriságú típusok lehetnek szabályos képzésűek is, és ebben az esetben erősítik az adott képző produktivitását. (Ehhez a kérdéshez bővebben lásd a 6. pontot.)

A kvantitatív produktivitási mutatók célja a fentiek alapján nem diszjunkt besorolás a „produktív” vagy „nem produktív” kategóriába, értelmezésükkel kapcsolatban mégis felmerül a tertium comparationis kérdése: mihez képest nevezhető egy konstrukciós séma inkább vagy kevésbé produktívnak? Baayen (1992) bevezet egy produktivitási „nullpontot”: a per definitionem terméketlen szimplexek (vagyis nem képzett szavak) produktivitási mutatói alkotják a viszonyítási alapot az általa alkalmazott kvantitatív módszerekben, nem produktívnak azokat a képzőket nevezi, amelyek mutatói nem haladják meg szignifikánsan a szimplexek értékeit (Baayen 1992: 116). Tekintve, hogy Baayen vizsgálatának tárgya a főnévképzés, a jelen tanulmány viszont igei konstrukciókra vonatkozik, a szimplexek vizsgálata itt kevésbé lett volna célravezető. A szimplex igeik hapaxai gyakorlatilag mind pusztán elemzési hibákból adódnak, ellenben számos nagy gyakoriságú szimplex ige van a magyarban (például *van, jön, megy*), ezekből az adatokból a potenciális produktivás gyakorlatilag nulla; a típusgyakoriság megállapítását a számos elemzési hiba jelentősen megnehezíti.

Az abszolút adatoknál sokkal érdekesebb számunkra az egyes passzív jelentésű konstrukciók produktivitási mutatóinak összevetése, szem előtt tartva, hogy a produktivás graduális fogalom; az egyes sémák produktivitási mértékének arányát össze tudjuk hasonlítani. Ezenkívül összehasonlítani alapul szolgálnak számunkra egy analitikus konstrukció, a *V-ni fog* konstrukció produktivitási mutatói. Az analitikus „kontroll”-konstrukciót egyrészt azért vontuk be, hogy teszteljük, befolyással van-e a konstrukció analitikus jellege a produktivitási mutatókra pusztán a nagyobb szintaktikai kombinálhatóság révén. Másrészt a konstrukció a deriváció-inflexió kontinuumon inkább az inflexió tartományban helyezkedik el, melyre többek között magasabb fokú produktivás jellemző (ezt Gaeta 2007 korpuszvizsgálata szabályos szintetikus elemekkel kapcsolatban igazolta).

## 2.1. Kvantitatív produktivitási vizsgálatok

A produktivitás kvantitatív vizsgálatának számos módszere van, ezek közül a globális produktivitás csak egy – ez azonban az egyik leggyakrabban használt, elméleti szempontból jól alátámasztható módszer (lásd fent), mely statisztikailag is az egyik legjobb közelítésnek számít (vö. Zeldes 2012). Zeldes számos kvantitatív produktivitási vizsgálati módszer összehasonlítása során azt találta, hogy két, többdimenziós módszer tükrözi legjobban a nyelvi produktivitást: a szókészlet-növekedési görbe (vocabulary growth curve, VGC; ez a folyamatos adatfeldolgozással nyert görbe azt mutatja, hogyan növekszik a korpusz nagyságával az adott sémával képezett konstrukciók száma), valamint a globális produktivitás ( $P^*$ ). Mivel Baayen (1992) alapján tudjuk, hogy ez a kettő szorosan korrelál egymással – a VGC görbe érintőjének meredekségét adja az  $n1/N$  képlet, azaz a sémával képzett hapaxok száma, osztva a sémával képezett tokenek számával (lásd Baayen 1992: 115) –, számunkra elegendőnek bizonyult a kétdimenziós  $P^*$  meghatározása a vizsgált konstrukciókra.

A globális produktivitás ( $P^*$ ) két dimenzióját a realizált/megvalósult produktivitás és a potenciális produktivitás adja. A realizált produktivitás egyenlő az adott sémával képezett típusok számával egy adott korpuszban – a továbbiakban egyszerűen típusgyakorisággként hivatkozunk rá –, míg a potenciális produktivitást úgy kapjuk meg, hogy a sémával létrehozott hapaxok számát elosztjuk a sémával létrehozott tokenek számával. Az eredmény egy koordináta-rendszerben ábrázolható, ahol az origótól mért távolság mérvadó a produktivitás szempontjából. További, a produktivitás kvantitatív mérésére használatos módszerek például:

- a TTR (type-token-ratio, típus-példány arány), mely klasszikusan a lexikai változatosságot fejezi ki, de nem ad számot például a hapaxok, illetve neologizmusok előfordulásáról, tehát nem ad teljes képet a produktivitásról, amennyiben a produktivitás fogalmát az új szavak létrehozásának lehetőségeként definiáljuk;
- a (kétfajta) HTR, melyet hapax/token ratioként, azaz hapax-token arányként (Stefanowitsch 2020: 318), illetve hapax/type ratioként, azaz hapax-típus arányként (vö. Tanos 2012, van Wettere 2022) szoktak számitani;
- a szintén a hapaxok számával operáló „terjeszkedő” produktivitás (expanding productivity, a magyar terminushoz lásd Kalivoda 2019), ez az adott sémával képezett hapaxok aránya a korpuszban található összes

hapaxhoz; a számítási képletnek megfelelően annak a valószínűségét méri, hogy egy új képzésű szó, azaz hapax, éppen a vizsgált sémával valósul meg;

- a potenciális produktivitás, mely a globális produktivitás egyik dimenzióját képezi, számítása: a hapaxok száma osztva a sémával létrehozott tokenek számával; annak a valószínűségét vizsgálja, hogy egy adott sémával létrehozott példány új lexéma;
- más globális produktivitási index például a Baayen által alkalmazott S/V, illetve az Aronoff (1976) által alkalmazott V/S, ahol V a sémával létrehozott típusok számát, S a teljes szókészletet jelöli.

A kvantitatív produktivás-számítási módszerek átfogó ismertetését – az itt nem szereplő módszerekét is – lásd Baayen (2009), Zeldes (2012) és Hilpert (2013) műveiben.

## 2.2. A produktivitási mérések kiterjesztése szintaktikai konstrukciókra

A fent ismertetett kvantitatív produktivitási indexek elsősorban morfológiai produktivás mérésére lettek kidolgozva, joggal merülhet fel tehát a kérdés, hogy mennyiben alkalmazhatóak ezek a mutatók analitikus igealakok – szintaktikai jellegű – konstrukcióira. Ehhez a következőket érdemes figyelembe venni. Az általunk választott elméleti keret (használat alapú kognitív nyelvészet) nem feltételez lényegi különbséget a különböző komplexitású és absztrakciós szintű nyelvi jelek működési elvében, ezért – ahogy Stefanowitsch (2006) is megállapítja – nemcsak morfológiai jellegű konstrukciókra alkalmazható Baayen indexe. Stefanowitsch esettanulmánya is egy grammatikai jellegű szó szerkezetet vizsgál (a német *haben zu* + Infinitiv konstrukciót). Zeldes (2012) a lexikogrammatikai kontinuum elvére építve egy kifejezetten szintaktikai jellegű konstrukció típusra, az argumentum-szerkezetre alkalmazza a produktivitási vizsgálatokat. Tanulmányunk metodológiai újdonsága abban áll, hogy itt két-két, hagyományosan morfológiai, illetve szintaktikai jellegű<sup>8</sup> konstrukció produktivitásának összevetése a cél, melyek funkcionálisan hasonlóak egymáshoz, és egyaránt egy-egy igei komponens található bennük.

<sup>8</sup> Bár a hagyományos grammatika (például Keszler szerk. 2000: 27–28) használja a „morfológiai típusú szó szerkezet” terminust a segédigés szerkezetekre vonatkozólag, a *V-va/-ve van* típusú konstrukciót nem sorolja ezek közé. Imrényi (2017) ellenben e konstrukciókat olyan komplex magmondattípusnak tekinti, amelynek elemei együtt reprezentálják a lehorgonyzott folyamatot, valamint megemlíti passzív funkcionális jellegüket is.



### 3. A vizsgált konstrukciók szemantikai jellemzése

#### 3.1. Konstrukciós poliszémia

A természetes nyelvek konstrukcióinak egyik alapvető tulajdonsága a poliszémia, nincs ez másként a passzív kifejezőeszközök esetében sem. A funkcionális és kognitív elméletek fokozott figyelmet fordítanak a poliszémia modellálására a szemantikus („grammatikai”) jelentésű konstrukciók esetében is (például Janda 1990). A passzív jelentésű konstrukciók szemantikai leírásához az elméletet Modrián-Horváth (2024) tárgyalja részletesen, a jelen tanulmányban csak nagy vonalakban vázoljuk a konstrukciók által lefedett fogalmi tartományokat.

A vizsgált két szintetikus és két analitikus kifejezőeszköz is, eltérő mértékben, a passzívon kívül más, például mediális, rezultatív vagy akár progresszív jelentéstartalmak megkonstruálására is szolgál (Modrián-Horváth 2024). Ez a *V-ódik/-ődik* és a *V-va/-ve van/lesz* konstrukciók esetében nyilvánvalóbbnak tűnik, mint a par excellence passzívként számontartott *V-(t)atik/-(t)etik* sémával kapcsolatban. Azonban, ahogy H. Tóth (1991) Bécsi Kódexen végzett korpuszvizsgálata kimutatta, a *-(t)atik/-(t)etik* képző is számos jelentésben volt használatos (passzív, mediális, bennható, reflexív, kiható), melyek közül H. Tóth szerint a passzív jelentés produktívan (elavult, régies használatban), a mediális jelentés pedig nem produktívan jelen van a mai magyar nyelvben. A *V-ódik/-ődik* konstrukciós sémával kapcsolatban Szili (1999: 356) megállapítja, „hogy az *-ódik*, *-ődik* a kifejezendő szituációktól függően íveli át a passzívtól a mediálisig terjedő rokon jelentéstartományokat”.

A *V-va/-ve van* konstrukció esetében Honti–H. Varga (2006/2013) alapján megállapíthatjuk, hogy nagyfokú poliszémia jellemzi ezt a kifejezőeszközt. Honti és H. Varga (2013) Szepesy (1972) alapján a konstrukció következő funkcióit sorolják fel: állapotjelölés, állapotváltozás; befejezettség, eredmény; személytelenség; a cselekvés intenzitásának a kiemelése; tömörítő árnyalás; a cselekvés befejezettségének kifejezése; a mozgás befejezettségének a jelölése; testhelyzet, biológiai állapot szemléltetése; állandósult kifejezések (Honti–H. Varga 2013: 260).

A jelen tanulmány a konstrukciós poliszémiát a konstrukciók egyik alapvető tulajdonságaként kezeli, melyre jellemző a radiális jelentésszerkezet (vö. Lakoff 1987, Lewandowska-Tomaszczyk 2007, grammatikai vonatkozásban Janda 1990), valamint a családi hasonlóság. Közös jelentésnyalábok figyelhetők meg például a *V-va/-ve van* konstrukció használataiban: passzív és

rezultatív például a *be van zárva* jelentése, reflexív/mediális-rezultatív a *meg van borotválkozva*, míg passzív-imperfektív/progresszív az *éreztetve van*. Bár az empirikus elemzésnél ez operacionalizálási nehézséget okoz, mégis természetes, hogy az egyes jelentések – különösen az olyan absztrakt szemantikai dimenziókban, mint a mediális és a passzív jelentés – nem különülnek el egymástól élesen (vö. 3.2.; a passzív és mediális igenem fokozati jellegéhez lásd Tolcsvai Nagy 2017: 331).

### 3.2. Passzív és mediális igenem: elhatárolási nehézségek

Egy hármas igenemfelosztás (aktív, mediális,<sup>9</sup> passzív) esetén a passzívum határterülete a mediális igenemmel érintkezik, itt merülnek fel elhatárolási nehézségek az elemzésnél.

Tolcsvai Nagy Gábor (2017: 334) jellemzése szerint a prototipikus szenvedő (passzív) igenem trajektora páciens, „a folyamat rajta megy végbe, nyíltan jelen van, a mondatban kidolgozást kíván”, a landmark „erőforrás, rejtett (implicit) cselekvő, amely az eseményt végrehajtja”. A meghatározás szerint „[a] szenvedő igében az aktív cselekvő figurája háttérbe kerül, a folyamatot elszenvedő figura előtérbe helyeződik (a cselekvő ige jelentésszerkezetének fordított változataként)”. A Tolcsvai Nagy szerkesztette Nyelvtan (ONyt.) morfológiai fejezetében Ladányi (2017) említi az alacsony gyakoriságú passzívképzőt *-(t)atik/-(t)etik*, míg a mondattani fejezetben Imrényi (2017: 688k.) említést tesz az analitikus *V-va/-ve van/lesz* konstrukcióról a „régies” szintetikus passzív „ma is általános [...] típusú” alternatívájaként. Imrényi (2017: 689) szerint a „passzív [...] igék (például *becsukatik*) abban különböznek a szintén Páciens alanyú ún. mediális igéktől (például *becsukódik*), hogy a passzív igék esetében a konstruálás része az Ágens is (vö. *Az ajtó magától becsukódott. \*Az ajtó magától becsukatott*). Ez az okozó szerepű, **aktív, tudatos szereplő** azonban 'nincs profilban' [...] [kiemelés a szerzőktől]. Az ONyt. tehát a **prototipikus**<sup>10</sup> szenvedő ige leírására törekszik, amely egy **intencionális cselekvést** és cselekvőt involvál.

A mediális igék jelentésszerkezetében ezzel szemben a trajektor végpont szerepű, experiensként, témaként vagy elszenvedőként konstruált figura,

<sup>9</sup> Az ONyt.-ban szereplő felosztásból (lásd Tolcsvai Nagy 2017) a visszaható és kölcsönös igéket jelentéstől függően, Kemmer (1993) alapján rokoníthatjuk a mediális igékkel, az elemzés során a vizsgált konstrukciókban adatolt visszaható igéket a mediális csoportba soroltuk (például *meg van borotválkozva*). A műveltető igék a cselekvőkkel rokoníthatók.

<sup>10</sup> A prototípus elvű kategorizációhoz vö. Rosch (1978).

„amely vagy aki valamilyen történésnek, állapotváltozásnak, állapotnak az elsődleges szereplője. A másik figura kezdeményező szerepű, energiaforrásként konceptualizálva” (Tolcsvai Nagy 2017: 332). A landmark többnyire nem nyílt, de fokozati különbségek figyelhetők meg a mediális igék között arra nézve (is), hogy „a kezdeményező szerepű figura mennyire cselekvő és mennyire nyílt a jelenléte az ige jelentésszerkezetében” (lásd uott). A mediális igék jellemzése során Tolcsvai Nagy (2017: 333) megállapítja, hogy egy részük (ilyen például a *leng* ige) a szenvedőkhöz áll közel; a passzív és a mediális határán van „a *bevágódik* (az ajtó a huzattól), *borul* (az építmény)”. Egy másik kognitív szempontú csoportosításban Laczkó Krisztina (Laczkó 2023: 79) leginkább Kemmer (1993) nyomán, de E. Abaffy (1978) felosztásával is kompatibilis módon, azokat az igéket, melyeket „a **teljes szándékolatlanság** és az eseményszerkezet és/vagy szereplők kifejtetlenségének legerősebb foka” jellemez, passzív mediálisoknak nevezi;<sup>11</sup> ide tartozik a *letörik* és a *kiderül* ige (lásd uott, kiemelés a szerzőktől).

A passzív és mediális igék elhatárolását a legtöbb szerző a szándékosság/szándékolatlanság mentén próbálja elvégezni: a prototipikus passzív (szenvédő) ige egy cselekvés fókuszváltással létrejött megkonstruálása, a cselekvés esszenciális összetevője egy intencionális ágens, mely landmarkként konstruálódik meg (a közvetlen hatókörben van, azaz a jelenet szereplőjeként akkor is feldolgozzuk, ha nincs explikálva), míg a mediális iginél nem található ilyen. Az intencionalitás kívánalma más megközelítésekben is meghatározó szerepet játszik, így például Siewierska (2013) szándékosságot kifejező határozókkal teszteli a passzív konstrukciókat a mediális (antikauzatív) konstrukcióktól való elhatárolás során. Ennek az alapján az (1) mondat passzív, míg a (2) mondat antikauzatív minősítést kap.

- (1) The door was opened deliberately.
- (2) The door opened (\*deliberately).

A tesztelés a jelen tanulmány keretében metodológiailag nem megoldható, ellenben a tényleges nyelvi tevékenységben vizsgálható.<sup>12</sup> A „szándékosan

<sup>11</sup> Vö. Lengyel (2000a: 86 és 2000b: 69) mediopasszív kategóriáját és E. Abaffy (1978) mediális igéinek első két csoportját is.

<sup>12</sup> Hasonlóan a tesztelés elkerülése céljából Modrián-Horváth (2024) bemutatja, hogy az MNSZ2 adatai szerint mind a négy itt vizsgált konstrukció használatos az által névutós kifejezéssel, noha a szakirodalom szerint ez csak a *V-(t)atik/-(t)etik* esetében lehetséges; például „a hatóságok által volt beépítve” (doc#406), „Anne Frank keze által íródott” (doc#1057).

megrongálódott” szókapcsolatra a Google kereső segítségével hat különböző találatot kaptunk (lásd (3)–(4)).

- (3) Legkésőbb szeptember vége óta az Oroszországból Németországba irányuló gázz szállítások tömegesen akadályoztatva vannak, miután az Északi Áramlat 1 és 2, az Oroszország és Németország közötti két fő gázvezeték feltehetően *szándékosan megrongálódott* szeptember 26-án.<sup>13</sup>
- (4) Ekkor a festmény *szándékosan megrongálódott*.<sup>14</sup>

A *szándékosan/direkt elfelejtődött* kifejezésekre is találunk példákat (lásd (5)–(6)).

- (5) Éppen ellenkezőleg: ebben az esetben azt gondolom, hogy ami elveszett, vagy *szándékosan elfelejtődött*, az éppen a cseh filmművészet identitásának és sajátosságának lényegéhez tartozhat.<sup>15</sup>
- (6) Ja, mintha *direkt elfelejtődött* volna..., mint az örmény nyelvtudásom...<sup>16</sup>

Azonban a *szándékosan*-teszt véleményünk szerint mégsem árul el sokat a cselekvés intencionalitásáról, már csak azért sem, mert különböző kontextusokban a *szándékosság* foka eltérő lehet, tehát egyazon igének létezhet mediális és aktív vagy mediális és passzív olvasata is. Ezenfelül a konstruálásban nagyfokú metapragmatikai tudatosság is szerephez juthat, ironikus olvasatot indukálva (vö. (5)–(6)) azáltal, hogy az aktív perspektívához kötődő *szándékosan* a mediopasszív perspektívájú ige által előhívott folyamat körülményét dolgozza ki.<sup>17</sup>

A passzív és mediális szemantikai tartomány elhatárolása könnyebb, ha az alapbeállítás szerinti trajektor (legtöbbször ágens, néha experiens)

<sup>13</sup> [https://forum.index.hu/Search/showArticleResult?topic\\_id=9136261&aq\\_ext=1&aq\\_text=megrong%E1l%F3dott](https://forum.index.hu/Search/showArticleResult?topic_id=9136261&aq_ext=1&aq_text=megrong%E1l%F3dott)

<sup>14</sup> <https://hu.blogotirni.com/articles/elado-vaszil-iliev-utolso-portreja-bulgaria-today.html> [2023. 07. 27.]

<sup>15</sup> <https://www.apertura.hu/2014/tel/hanakova-filmek-melyeket-szegyellunk/> [2023. 07. 25.]

<sup>16</sup> <https://holdkatlan.hu/index.php/szepirodalom/proza/9392-piro-m-peter-mozgo-unnep> [2023. 07. 25.]

<sup>17</sup> Köszönjük, hogy Palágyi László e körülményre felhívta a figyelmünket lektori véleményében.

a fókuszváltó konstrukcióban is a közvetlen hatókörben marad (vö. Imrényi 2017: 688–689).

Gyakran megfigyelhető azonban a passzív mintázatok metonimikus és metaforikus használata. A prototipikusnak tekinthető, intencionális (humán) ágens által végzett cselekvés sokszor valamilyen eszköz közvetítésével irányul az elszenvedőre, az oksági láncolat (causal chain; Croft 1991, 2001) alapján; az ágensi „behatás” itt az eszközhasználatától (például *átolajozódik*) akár egyetlen kattintásig terjed (egy tartalom *letöltődik*). A ráhatottság egyértelműen megvalósul, de a cselekvés kivitelezője ezekben az esetekben nem határolódik körül egyértelműen. Nem prototipikusan, de passzív jellegű a konstruálás például nem ágensi okozó esetében is (*elfelejtődik*); míg sok nem konkrét folyamat konstruálása (például *konstruálódik*, *elsáncolódik*) már szintén a mediális tartomány irányába mutat, mert az energiaforrás ezekben az esetekben nehezen azonosítható.

A passzív jelentéseknek a mediálistól való elhatárolása tekintetében érdemes egy pillantást vetni az „elismerten” grammatikalizálódott passzív konstrukciókkal rendelkező nyelvekre is. Az angol passzív használata mellett megjelenő, illetve kitehető *by*-konstrukciót szimptomatikus értékűnek szokták tekinteni, mely azt bizonyítja, hogy „valódi” passzívról és nem más predikatív konstrukcióról van szó (vö. például Siewierska 1984). Azonban rendszeresen találunk olyan megvalósulásokat, amelyeknél nem (intencionális) ágensi okozó mellett fordul elő a passzív forma, az angol *by*-frázis nem prototipikus ágens<sup>18</sup> fejez ki. Az alábbi mondatpárok közül az első mondat mediopasszív (lásd (7), (9)), a második ágenst tartalmazó mondat (lásd (8), (10)).

(7) Last weekend’s hiking trip was ruined *by the weather*.<sup>19</sup>

(8) I keep having dreams about the people who hurt me on my birthdays, and my birthday was ruined *by them*.<sup>20</sup>

<sup>18</sup> Az ágensi dimenziók differenciált megközelítéséhez lásd Dowty (1991) protoágens koncepcióját.

<sup>19</sup> [https://www.mmpublications.com/kazakhstan/site/Content/FlipBooks/FBKazGrade8\\_sb/files/basic-html/page140.html](https://www.mmpublications.com/kazakhstan/site/Content/FlipBooks/FBKazGrade8_sb/files/basic-html/page140.html) [2023. 08. 17.]

<sup>20</sup> <https://www.quora.com/I-keep-having-dreams-about-the-people-who-hurt-me-on-my-birthdays-and-my-birthday-was-ruined-by-them-I-always-wake-up-panicking-and-full-of-guilt-Is-that-a-sign> [2023. 08. 17.]

- (9) You have already been physically injured *by the accident*<sup>21</sup>  
 (10) A second boy was shot and injured *by soldiers* in a nearby village<sup>22</sup>

A pontosabb szemantikai besorolásához fontos, a használatalapúságból következő alapelv a konstrukcióknak támogató közegükben való vizsgálata, így többek között a diskurzus szerepének, a kontextusnak és a nyelvi variációnak a figyelembevétele (vö. Ladányi–Tolcsvai Nagy 2008: 33). A nagy korpuszok használata során ez a követelmény (például a produktivitás mutatóinak megállapításánál) nem vagy nehezen kivitelezhető. Kézi elemzéssel azonban kisebb adathalmazon elvégezhető, a jelen dolgozatban is az egyes példányok – hapaxok – elemzése támogató közegben történt, ezáltal az egyes konstrukciós példányoknak a passzív-mediális-aktív kontinuumon való besorolása pontosítható volt. A *fűrészselődik* szó mint típus nem árul el semmit arról, hogy milyen ágens/ráható erő által történik a fűrészelés, ezáltal nehézségekbe ütközik a szó besorolása a mediális vagy passzív kategóriába; míg az alábbi kontextus figyelembevételével mediopasszív kategóriába sorolható az előfordulás (11):

- (11) Aztán amikor mélyebbre fűrészselődött a völgy, a barlang teljesen szárazzá vált. (doc#2573)

### 3.3. A V-(t)atik/-(t)etik konstrukció

A *-(t)atik/-(t)etik* képzőt a hagyományos nyelvtan az egyetlen, produktív passzívképzőként tartja számon (vö. Keszler szerk. 2000). Kenesei et al. (1998) korlátozottan tartja a képző produktivitását, azonban az esetleges alternatív passzív konstrukció tekintetében nem foglal állást. Ladányi (2017: 610) szerint a képző elvileg korlátozás nélkül használható minden tárgyias igével, azonban az egyes ide tartozó igék példánygyakorisága alacsony, és használata a hivatali nyelvbe, valamint a szépirodalom nyelvbe szorult vissza.

<sup>21</sup> <https://donaghuelabrum.com/2022/02/17/post-car-accident-checklist-for-pennsylvania-drivers/> [2023. 08. 11.]

<sup>22</sup> <https://www.reuters.com/article/uk-israel-palestinians-violence-idUKBRE90E0F020130115> [2023. 08. 11.]

### 3.4. A V-ódik/-ődik konstrukció

A V-ódik/-ődik konstrukciót számos leírás a szintén szintetikus V-(t)atík/-(t)etik séma potenciális utódjaként említi meg (lásd Szili 1999, Grétsy 2014, Ladányi 2017), ez a séma vehetné át „a” passzívképző szerepét a magyar nyelvben. Emellett szólhatna az is, hogy a konstrukciót produktívan használják nyelvjárási beszélők passzívképzőként a Tisza, Szamos környékén – például a *fa kivágódik* (Grétsy 2014). Amennyiben ez bekövetkezne, a következő, egy gazdasági szakkönyvből származó, a fák oltásáról szóló idézet félkövérrel kiemelt igéi nem tűnnének idegenszerűnek az olvasó számára (Tsötönyi 1831: 131, kiemelés a szerzőktől).

- (12) „A’ fa, vagy ág, mellybe oltanak, egyenesen kerekre *fűrészselődik*’s *faragódik*. Éppen a’ székén, velején, vagy közepén keresztül, ha lehet tsak annyira *hasitódik* be, a’ mennyire az oltó ág ék formára faragott vége le ér. Hogy a’ hasítás öszve ne menjen, tömpe vastag ék *verődik bele*. Az oltó ág gyűrűjén alól, két kis uj szélességre *vágódik el*, úgy *faragódik* laposra és élesre, hogy a bele vagy közepe meg nem *sértődik*.”

### 3.5. A V-va/-ve lesz és van konstrukciók

A két konstrukciót a *van* és *lesz* tövek szuppletív használata (például *lesz* mint a létige jövő ideje, *lenne/lett volna* mint a létige feltételes módú alakjai, *legyen* mint felszólító-kötő mód) miatt nem lehet egyértelműen elhatárolni egymástól, egyetlen konstrukcióként való kezelésük mégsem elfogadható. Ennek oka leginkább abban rejlik, hogy a *van* statikus, a *lesz* dinamikus sematikus jelentéssel bír. Nehéz lenne kijelenteni, hogy az *át lesz irányítva* konstrukció kizárólag jövő időben horgonyzódik le, időjelentése inkább aktuális vagy általános jelen vonatkozású (míg más előfordulások természetesen a jövőre vonatkoznak). Továbbá a V-va/-ve *volt* mellett az elkülönült V-va/-ve *lett* múlt idő is az önálló konstrukcióként való vizsgálatot támogatja. Elővizsgálataink szerint (és ezt a jelen kutatás is igazolja) a V-va/-ve *van* konstrukció magasabb fokú poliszémiát mutat a V-va/-ve *lesz* konstrukciónál, valamint a passzív/mediális konstrukciós jelentés megoszlása tekintetében is erőteljes különbségek tételezhetők fel a két séma produktív használatában. Az ONYt. szerint (Imrényi 2017: 688–689) az *Ajtó be lett csukva* típusú konstrukció a „régies” szintetikus passzív „ma is általános [...] típusú” alternatívája

(Imrényi 2017: 688k.). A *V*-*va/-ve van/lesz* konstrukciókat Kádár–Németh (2009) szerint produktívan használják a moldvai magyar nyelvjárásokban.

### 3.6. További passzív jelentésű magmondat-konstrukciók

Meg kell említeni, hogy a nem modális (medio)passzív tartományban való konstruálásra további (magmondat-) konstrukciók is alkalmasak, így például a befejezett melléknévi igeneves magmondatok (például *a torony felállványozott, a gázellátás biztosított*), vagy más, alapvetően mediális jelentésű sémák használata (például a *V-ít* tranzitív sémával kontrasztot képező *V-ul/-ül* használata a (13–15) példákban).

- (13) Várjon, amíg a frissítés *települ!* (Adobe Reader rendszerüzenet, 2019. márc.)
- (14) A Microsoft Edge *bezárult*, amíg néhány lap meg volt nyitva. (Microsoft rendszerüzenet, 2023. 07. 21.)
- (15) Híreink pedig folyamatosan *frissülnek* az interneten (Kossuth rádió, Déli krónika, 2019. márc. 9., 12.39)

Ezeket a konstruálási sémákat jelen korpuszelemzés nem vizsgálja.

## 4. A konstrukciók globális produktivitási vizsgálata az MNSZ2-UD korpuszban

Az első eredményeink ismertetése előtt nagy vonalakban bemutatjuk a gépi módszerek alkalmazásának előnyeit és hátrányait azzal a céllal, hogy a kutatás lehetőségeit és korlátait kontextusba helyezzük. A gépi módszer előnyei között elsőként azt említenénk, hogy ezáltal rövid idő alatt hatalmas mennyiségű adat elemezhető, és olyan tendenciák rajzolódnak ki, amelyeket hagyományos módszerrel nem lehetne elérni. Másodsor, több műfajt képes magába foglalni, mint a kézzel végezhető elemzés. Stefanowitsch (2006) minél nagyobb, minél több műfajt tartalmazó korpusz használatát javasolja. Ennek kiemelt jelentősége van a produktivitás vizsgálatokor, hiszen Baayen (2009) figyelmeztet arra, hogy a produktivitás igen eltérő lehet az egyes műfajokban (sőt egyedi különbségeket is mutat az egyes szerzők repertoárjában). Harmadszor, az elemzéshez használt forrásanyagok, szoftverek, programkódok birtokában a teljes elemzési folyamat reprodukálható.



A kapott eredmények könnyebben ellenőrizhetők mások által is, és könnyebben felhasználhatók későbbi kutatásokban. Hátrányként pedig azt kell elkönyvelnünk, hogy a gépi elemzés minősége nem éri el egy kézzel összeállított, átnézett adathalmaz minőségét.

#### 4.1. A felhasznált korpusz

A Magyar Nemzeti Szövegtár 2 (Oravecz et al. 2014) egy 1,04 milliárd szövegszóból álló, automatikusan elemzett, általános célú korpusz, amely a 20. század végének és a 21. század elejének magyar nyelvhasználatának reprezentálására törekszik. A kutatásunkban ennek a legújabb – megjelenés előtt álló – változatát vettük alapul, az MNSZ2-UD korpuszt (Kalivoda et al. 2024). Ez a verzió három szempontból tér el az eredeti MNSZ2-től. Egyrészt számos korpusztisztítási lépésen ment keresztül: ilyen például a bekezdésszintű duplumszűrés, a nem magyar nyelvű bekezdések automatikus azonosítása és törlése, valamint az extrém hosszú mondatok kiszűrése. Ezáltal a szöveg minősége jobb, de a korpusz mérete kisebb az MNSZ2-höz képest (776,9 millió szövegszó). Másrészt a stílusregiszter és a régió mellett elérhetővé vált a publikálás évére vonatkozó metaadat is a szövegek 93,8%-a esetében. Harmadrészt a korpusz új elemzési szintekkel gazdagodott: az emMorph (Novák et al. 2016) által létrehozott, részletes morfológiai elemzéssel, valamint függőségi elemzéssel, amely a Universal Dependencies formalizmusát követi (de Marneffe et al. 2021).

A jelenlegi kutatáshoz elsősorban arra volt szükség, hogy a lehető legpontosabban tudjuk azonosítani az igét, amelyhez egy-egy elvált igekötő tartozik. Ezért úgy döntöttünk, hogy a korpusz szövegeit végigelemezzük az emPreverb eszközzel is (Pethő et al. 2022). Ez a szabályalapú algoritmus kifejezetten a magyar igék és igekötők összekapcsolására készült, és ezzel érhető el a legpontosabb annotáció Pethő et al. (2022: 86–87) kiértékelése alapján.

#### 4.2. Az adatgyűjtés folyamata

A teljes korpuszból lekérdeztünk minden olyan szóalakot, amely számunkra releváns képzőt tartalmaz, tehát a *-va/-ve*, *-atik/-etik/-tatic/-tetik* vagy *-ódik/-ődik* szuffixumok valamelyike azonosítható benne. Minden szóalak esetében megvizsgáltuk, hogy tartozik-e hozzá prefixált vagy elvált igekötő. Az utóbbit az emPreverb annotáció segítségével találtuk meg. Ennek a lépésnek az volt a célja, hogy az igekötős igelemmák egyforma reprezentációt

kapjanak, függetlenül az igekötő pozíciójától (például az *elkezdődött* és a *kezdődjön el* lemmája egységesen *el#kezdődik* lett). Egymás alakvariánsainak tekintettük és azonos alakra normalizáltuk a következő igekötőket: *belé* → *bele*, *bé* → *be*, *föl* → *fel*, *föül* → *felül*, *fönn* → *fenn*, *odább* → *odébb*, *reá* → *rá*, *teli* → *tele*.

A *-va/-ve* képzős alakok esetében megvizsgáltuk azt is, hogy létigével állnak-e (amelyet maximum három tokenes bal és jobb kontextusban kerestünk), és ha igen, pontosan milyen alakban áll a létige. A 2.5. pontban megfogalmazottak alapján arra számítottunk, hogy a *V-va/-ve + lesz*, illetve a *V-va/-ve + van* szerkezetek használata, ezáltal a produktivitása is számottevő eltéréseket mutathat. A korpuszban különálló lemmaként jelenik meg a *lehet*, amelyet *lesz-re*, valamint a *nincs* és a *sincs*, amelyeket *van-ra* konvertáltunk át. Az igekötős *-va/-ve* képzős alak jobb kontextusában megjelenő létige kapcsán azzal szembesültünk, hogy szinte minden találat fals pozitív, azaz nem a keresett konstrukció egy példánya (például a *szavazatok közel 53%-át megszerezve lett a Francia Köztársaság XXIII. elnöke*, ahol a *megszerezve lett* hibás pozitív találat a keresett konstrukció szempontjából). Ezért a vizsgálatunkhoz csak azokat a találatokat őriztük meg, ahol a létige megelőzi a *-va/-ve* képzős alakot (például *lett megszerezve*, *meg lett szerezve*), vagy ahol a létige előtt álló *-va/-ve* képzős alak igekötő nélküli határozói igenév (például *tájékoztatva lett*).

Minden találat esetében megőriztük a teljes mondatot, valamint a publikálás évére és a dokumentum regiszterére vonatkozó metaadatokat. A kvantitatív produktivitási vizsgálatunkban csak azokat a találatokat vettük figyelembe, amelyek 2000-ből vagy azt követő évekből származnak.

A bemutatott adatgyűjtési folyamatot elvégeztük a *V-ni fog* konstrukció példányainak azonosítására is. Itt a (ragozott és ragozatlan) infinitívuszi szóalakot kerestük elsőként, majd azonosítottuk ennek a lehetséges igekötőjét, és maximum három tokenes bal és jobb kontextusban kerestük a *fog* lemmát. Ezután elvégeztük a passzív jelentésű konstrukciók, valamint kontrollcsoportként a *V-ni fog* szerkezet globális produktivitásának vizsgálatát, egyrészt a teljes korpuszra vetítve, másrészt az egyes regiszterek szerinti bontásban. A következő két fejezetben ismertetjük a méréseink eredményét.

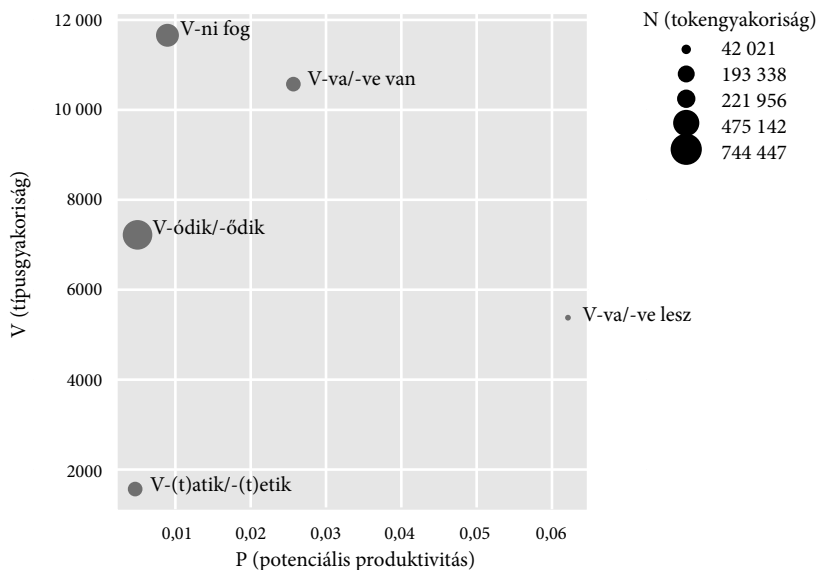
### 4.3. Az egész korpuszra vetített eredmények

A globális produktivitás megállapításához négy értékre van szükség minden vizsgált konstrukció esetében. Ezek a tokengyakoriság (hány példányban adatható a szerkezet a korpuszban), a típusgyakoriság (hány igelemméval jelenik meg), a hapaxok száma (hány egyszer előforduló igelemméval szerepel), valamint a potenciális produktivitás, amelyet úgy kapunk, hogy a hapaxok számát elosztjuk a tokengyakorisággal.

1. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n<sub>1</sub>) és potenciális produktivitása (P) a teljes korpuszban

<b>szerkezet</b>	<b>N</b>	<b>V</b>	<b>n<sub>1</sub></b>	<b>P</b>
V-ódik/-ődik	744 447	7203	3647	0,0049
V-va/-ve van	193 338	10 539	4978	0,0257
V-(t)atik/-(t)etik	221 956	1528	1033	0,0047
V-va/-ve lesz	42 021	5341	2616	0,0623
V-ni fog	475 142	11 625	4202	0,0088

A globális produktivitást úgy kapjuk meg, hogy a potenciális produktivitást és a típusgyakoriságot egy-egy dimenzióknak tekintjük, ezáltal kétdimenziós térben ábrázolni tudjuk (az előbbit egy koordinátarendszer X-tengelyén, az utóbbit az Y-tengelyén szokás megjeleníteni). A vizsgált konstrukciók egy-egy adatpontként elhelyezhetők ebben a térben az értékeiknek megfelelően. Még összetettebb képet kaphatunk az adatról akkor, ha az egyes konstrukciók tokengyakoriságát az adatpontok nagyságával szemléltetjük.



1. ábra. A vizsgált konstrukciók globális produktivitása a teljes korpuszban

Az 1. ábráról a következők olvashatók le:

- A *V-(t)atik/-(t)etik* képző a mai magyar nyelvben igen korlátozottan produktív: mind a potenciális produktivását, mind a típusgyakoriságát nézve alacsony értékeket látunk. Egyedül a tokengyakorisága számottevő (a 3. helyen végzett a vizsgált öt konstrukció között). Ennek a konstrukciónak egyúttal a legalacsonyabb a TTR-aránya (típus-példány aránya) is, mely nagyságrendi különbséget mutat a *V-va/-ve van* és *V-va/-ve lesz* sémákhoz képest (lásd 2. táblázat).
- A *V-ódik/-ődik* globális produktivitása már jelentősebb: típusgyakoriságát tekintve a középmezőnyben szerepel, de a potenciális produktivitása alacsony (a magas tokengyakoriságához képest kevés hapaxban jelenik meg).
- A vizsgált passzív jelentésű konstrukciók közül leginkább a *V-va/-ve van* és a *V-va/-ve lesz* produktív, de más-más dimenziók mentén: az előbbinek a típusgyakorisága, az utóbbinak a potenciális produktivitása a kiemelkedő.

- Potenciális produktivitás szempontjából 0,005 alatti értéket mutat a *V-ódik/-ődik* és a *V-(t)atik/-(t)etik*; 0,01 alatti a *V-ni fog*, melynél csaknem háromszor produktívabb a *V-va/-ve van*, és utóbbinál is csaknem két és félszer magasabb értéket mutat a *V-va/-ve lesz*.
- Az összehasonlítási pontként használt *V-ni fog* szerkezet a legnagyobb típusgyakoriságú, de az alacsony potenciális produktivitását tekintve inkább a szintetikus szerkezetekhez közelít.

2. táblázat. A vizsgált konstrukciók típus-példány aránya

szerkezet	N	V	V/N
<i>V-ódik/-ődik</i>	744 447	7203	0,0097
<i>V-va/-ve van</i>	193 338	10 539	0,0545
<i>V-(t)atik/-(t)etik</i>	221 956	1528	0,0069
<i>V-va/-ve lesz</i>	42 021	5341	0,1271
<i>V-ni fog</i>	475 142	11 625	0,0245

#### 4.4. Az alkorpuszok eredményei

A konstrukciók globális produktivitását megvizsgáltuk alkorpuszok, azaz regiszterek szerinti bontásban is. A következőkben csak azokat a megfigyeléseket emeljük ki, amelyek a teljes korpuszon végzett vizsgálathoz képest nagy eltérést mutatnak. Az egyes regiszterekre vonatkozó részletes eredményeket mellékletben közöljük.

- Szépirodalom (I. sz. melléklet): a teljes korpuszhoz nagyon hasonló eredményt kaptunk, csak annyi a lényegesebb eltérés, hogy a *V-va/-ve lesz* konstrukció típusgyakorisága alacsonyabb.
- Ezzel szemben a személyes alkorpusz (II. sz. melléklet) esetében a *V-va/-ve lesz* típusgyakorisága magas, felülmúlja a *V-ódik/-ődik* sémáét.
- A hivatalos alkorpuszban (III. sz. melléklet) a passzív jelentésű konstrukciók egymáshoz viszonyított távolsága hasonlóan alakul a teljes korpuszhoz; a *V-ni fog* szerkezet távolodik el minden másától a nagyobb típusgyakoriság irányába.
- A sajtónyelvi alkorpuszban (IV. sz. melléklet) megjelenő tendenciák egyeznek a hivatalos alkorpuszban látottakkal.

- Az eddigi mintáktól jelentősebben eltér a tudományos ismeretterjesztő alkorpusz (V. sz. melléklet): az összehasonlítási alapul használt *V-ni fog* szerkezet kisebb típusgyakoriságú, míg a *V-va/ve van* konstrukció feltűnően produktív mindkét dimenzió mentén.
- A beszélt nyelvi alkorpusz (VI. sz. melléklet) olyan szempontból kilóg a sorból, hogy valójában az „írott nyelvivel” állna kontrasztban, tehát a korábbi öt regiszterrel együttesen. A beszélt nyelvi alkorpusz a regisztert tekintve heterogén, informális és formális szövegeket egyaránt tartalmaz. A passzív jelentésű konstrukciók egymáshoz viszonyított globális produktivitása nem tér el jelentősen a teljes korpuszon mérttől. Talán azt lehetne kiemelni, hogy a *V-ni fog* szerkezet értékei alakultak máshogy (kimagasló típusgyakoriság és tokengyakoriság – a korábbi mérésekben rendszerint a *V-ódik/-ődik* tokengyakorisága volt a legmagasabb).

## 5. Kézi elemzéses vizsgálat a beszélt nyelvi alkorpuszon

Ahogy a 4. fejezet bevezetőjében említésre került, a gépi elemzésnek nemcsak hatalmas előnyei, de hátrányai is vannak. Gyakran fogalmazódik meg a gépi elemzéssel szembeni kritikaként, hogy az azt végző nyelvésznek nincs elég alapos betekintése, rálátása azokra az adatokra, amelyekkel dolgozik. Másrészt az automatikusan kinyert adatok között sok téves vagy irreleváns találat is lehet. Baayen (2009: 907) figyelmeztet a pusztán gépi elemzés „zajosságára”, illetve arra, hogy torzulhatnak az eredmények a gépi elemzésnél. A jelen tanulmány ezért egy korábbi, kézzel végzett elemzést is tartalmaz 3 sémával kapcsolatban (*V-va/-ve van*, *V-va/-ve lesz* és *V-(t)atik/- (t)etik*, amelyet az MNSZ2 (Oravec et al. 2014) egyik alkorpuszából, az ún. beszélt nyelvi alkorpuszból származó adatokon végeztünk (szemben a 4. fejezetben bemutatott, gépileg duplumszűrt változattal). A korpusz mérete mintegy 76,2 millió szövegszó, ez az MR1 Kossuth rádió bizonyos anyagait öleli fel az 2004–2012 évekből, felolvasott szöveget (hírek) és spontán beszélgetést (riportok) vegyesen tartalmaz. Az alkorpusz választását az indokolta, hogy várhatóan itt volt a legnagyobb a (fél)spontán megnyilatkozások aránya.<sup>23</sup> A negyedik, *V-ódik/-ődik* sémára nézve az adatokat kiegészítettük

<sup>23</sup> Ezenfelül véletlenszerű minták vizsgálata alapján ez volt a duplumokkal legkevésbé terhelt alkorpusz, szemben más alkorpuszokkal: a személyes alkorpuszban nagy mennyiségben fordul elő többszörösen beágyazott idézet (a forrásként szolgáló internetes fórumok egyik jellegzetessége), így ugyanaz a szövegdarab újra és újra megismétlődhet. Itt például egy

az MNSZ2-UD korpusz megfelelő alkorpuszán mért anyaggal, mivel ez kézi-  
leg feldolgozhatatlan mennyiségű előfordulást tartalmazott; a gépi adatokat  
csak összehasonlításképpen tüntettük fel és csillaggal jelöltük.

### 5.1. Az adattisztítás és -feldolgozás folyamata

Ahhoz, hogy a hapaxok pontos száma megállapítható legyen, szükséges volt  
a rendelkezésre álló szövegeket megtisztítani és elemezni.

Ennek első lépésében a kiválasztott grammatikai tulajdonságokkal rendel-  
kező lemmákat 300 karakteres kontextusukkal együtt lekértük az MNSZ2-  
ből. Az analitikus igealakok vizsgálatához a határozói igenévként annotált  
szavakat kerestük, az eredményt a *van/lesz* lemmával való együttes előfordu-  
lásra pozitívan szűrve egy mondaton belül.<sup>24</sup> Az igenévvél záruló elváló ige-  
kötős konstrukciók esetében (például *be van zárva*) 2, a nem elváló ige-kötős  
vagy ige-kötő nélküli esetben 3 szó beférkőzését engedjük meg (azaz az ige-kötő  
legfeljebb 4 szóra távolodott el az igenévtől, ami az ige-kötős konstrukciók ese-  
tében Kalivoda (2021) eredményei alapján a korpusztalálatok 99,88%-át lefedi  
– azzal a kitételrel, hogy a beszélt nyelvi korpusz esetében a találati arány né-  
mileg alacsonyabb lehet). A *van*, illetve *lesz* igével záruló konstrukciók esetén  
1 szó beférkőzését engedjük meg (például *zárva lesz*). A *V-(t)atik/-(t)etik* séma  
vizsgálatához *-atik*, *-etik*, *-tatik* vagy *-tetik* végződésű lemmákra kerestünk,  
majd negatívan szűrtük az eredményeket *-hatik*, illetve *-hetik* végződésre.<sup>25</sup>  
A *V-ódik/-ődik* séma kézi vizsgálata (kézi ige-kötőkapcsolás) a találatok magas  
száma miatt ebben a körben nem volt megoldható, ezért itt a korábbiakban  
bemutatott, automatikus ige-kötőkapcsolást tartalmazó részkorpuszt használ-  
tuk; itt is az *-ódik* vagy *-ődik* végződésű lemmákra kerestünk.

Az így kapott szövegeket Google Sheets táblázatokban és Python (v.3.10)  
programnyelvben dolgoztuk fel a következő lépésekben.

---

esetben (az *előadat* hibásan *előadatik*-ként lemmatizált szó esetében) egyetlen valós példány  
másolatai alkották csaknem az összes találatot, pontosan 260-at a 265-ből (a maradék öt  
találat között is van duplum, és mindegyik hibás lemmatizálásnak köszönhető).

<sup>24</sup> Keresőkifejezések: *V-va/-ve van* és *lesz* esetében inverz szórenddel: szóalak: határozói ige-  
név, pozitív szűrés [+1, +2] ablakban *lesz*, illetve *van* szótőre ige szófajjal, tetszőleges rago-  
zással, negatív szűrés [0, +1] ablakban tetszőleges írásjelre. Semleges szórenddel a pozitív  
és a negatív szűrési ablak egyaránt [-3, -1] volt, mert a felhasznált korpuszban az írásjelek  
nem számítanak külön tokennek.

<sup>25</sup> Keresőkifejezés: szótó: *\*(atik|etik|tatik|tetik)* és szófaj: ige, negatív szűrés 0 szó távolságban  
szótó: *\*(hatik|hetik)*

Elsőként a szövegeket kisbetűsítettük, eltávolítottuk az összes írásjelet (kivéve azokat, amelyekre technikai okokból szükség volt egy-egy esetben), és az így keletkezett változatokból töröltük a felesleges szóközlőket.

Az így egységesített szövegekben a következő szabállyal kerestünk duplikátumokat: ha két találati mintában legalább egy ötven karakter hosszúságú szakasz megegyezett, akkor az adott elemeket azonosnak tekintettük, és a második példányt töröltük. A viszonylag szigorú kritériumot az indokolta, hogy a rádiós szövegekből álló korpusz sokszor ugyanannak a diskurzusrészletnek több, részben eltérő lejegyzését tartalmazza, így bizonyos mértékű eltéréssel itt is ismétlődéseket találhatunk. Az ötven karakteres hátér bevallottan önkényes, azonban kézi ellenőrzéssel validáltuk több kisebb (≈3000 kontextus/minta) találatlistán, és nem produkált fals pozitív eredményeket. A (16a-b) szöveg példa az olyan minimális eltérésű ismétlődésre, ahol az egyetlen elhangzott diskurzusrészlet két különböző transzkripció miatt kétféleképpen került a korpuszba (például egyszer *szemész*, egyszer *zenész* alakban szerepelt ugyanaz a hangsor; az elütéseket a korpusz tartalmazza).

- (16) a. Tehát ez teljesen elképesztő, ugyanúgy működött mint ma, mondjuk például ismerek egy szemészt, aki ilyen reklámcsészeket készítettet és *a saját szemészeti pecsétje van beleütve*, az ő általa forgalmazott szemkenőcsöket akarta ezzel reklámozni. (doc#2608)  
 b. Ez teljesen elképesztő. Ugyanúgy működött, mint ma. Például, ismerünk egy zenészt, aki reklám csészeket készítettet, és *a saját pecsétje van beleütve*. Az ő által forgalmazott szemkenőcsöket akarta ezzel reklámozni. (doc#2647)

A duplikátumok eltávolítása jelentős változásokat hozott a fennmaradó szövegek mennyiségében, gyakran 25% körüli szövegről derült ki, hogy duplum. Ez a megfigyelés fontos kérdéseket vet fel a tisztítatlan korpuszokon végzett statisztikai elemzések megbízhatóságával kapcsolatban, amennyiben az eredeti korpusz nincsen valamilyen módon hatékony zajszűrővel ellátva.

Az így kapott találati listából kinyert hapaxok és alacsony tokengyakoriságú elemek ezután egyenkénti ellenőrzésre kerültek, kiszűrendő a fals találatokat, például bibliai vagy irodalmi idézeteket. A bizonyított, számunkra fellelhető forrásból származó idézeteket eltávolítottuk az adatsorból, mert sokszor többszáz éves vendégzsövegek kerültek volna be az adatok közé,



a vizsgálat viszont a mai magyar nyelvhasználatban tetten érhető produktivitást hivatott vizsgálni. A szállóigévé vált idézetek (például a *kocka el van vetve*) a listában maradtak. Ellenőrzésre szorult az is, hogy nem történt-e „véletlen” veszteség, amikor is egy szövegben két keresett kifejezés is szerepel, és az egyik alapján a szöveg duplumnak bizonyult, így törlődik, ami által a másik – releváns – találat elvész. Ezekben az esetekben a szöveget visszaállítottuk és a példányokat felhasználtuk a számításokhoz.

A gépi és kézi szűrés után fennmaradt példányok és hapaxok számát ezután határoztuk meg, és számoltuk ki a produktivitásra utaló értékeket.

A *V-ódik/-ődik* konstrukcióba tartozó megvalósulások nagy száma miatt (hozzávetőleg 106 000 token duplikátumok kiszűrése előtt és 76 000 utána) a kézi ellenőrzés nem volt lehetséges, így a duplikátumok eltávolítása után ezeknél csak az elveszett hapaxok visszaállítására került sor a produktivitási értékek kiszámítása előtt. Ebben a csoportban az igekötőkapcsolás is gépileg történt a beszélt nyelvi korpuszrészben (a módszerhez lásd 4.), ezért a *V-ódik/-ődik* sémára vonatkozó adatok csak korlátozottan hasonlíthatók össze a többi eredménnyel.

Az analitikus alakok és a *V-tatik/-tetik* séma vizsgálatánál kézi igekötőkapcsolás történt (utóbbinál ettől csak a legnagyobb, 100 fölötti gyakoriságú lemmák esetében tekintettünk el, mivel ezek az adatokba való betekintés alapján nem befolyásolták érdemben a produktivitási mutatót, például *születik, megszületik, adatik, megadatik*). A folyamat során az igekötős igéken kívül (például *be van zárva*) egy lemmához tartozónak tekintettük a következőket:

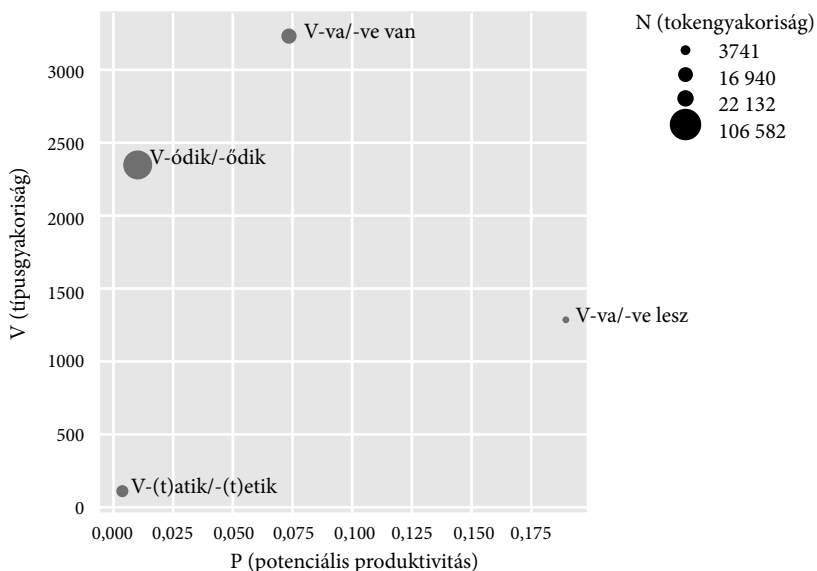
- Személyes névmások és igék kapcsolata (például *rám van bízva*), mely analóg a *rá van bízva* előfordulással, így egy lemmához soroltuk; egyekben az egységesség kedvéért a helyesírási előírásokat követtük (például a *számontart* és a *tart* külön lemmát képvisel, de a *férjhez megy* a *megy* lemmához tartozik).
- A regionális, stilisztikai alakváltozatokat, például *fel/föl, bocsát/bocsájt*, egy lemmához soroltuk, hogy ne gyarapítsuk indokolatlanul a hapaxok számát; hasonlóképpen az MNSZ2 az *alkuszik/alkud-* tőváltozatokat automatikusan egy lemmához sorolja.
- Néhány konstrukciónál még a kontextus pontos ismeretében sem lehetett megállapítani a bemeneti igetövet, ilyen például a *tör/török* igék *V-va/-ve van* alakja, itt az MNSZ2 lemmatizálását követve egy tőhöz soroltuk az adott igéket.

## 5.2. Eredmények

A globális produktivitás mutatóját itt is a 4.2. pontban bemutatott számítással mértük. A 3. táblázat az egyes sémák adatait, a 2. ábra ezek grafikus ábrázolását mutatja be.

3. táblázat. Globális produktivitás a beszélt nyelvi korpuszban, kézi elemzés  
(A \*-gal jelölt sor adatai automatikus elemzésből származnak)

szerkezet	N	V	$n_1$	P
V-ódik/-ődik*	106 582	2358	1094	0,0103
V-va/-ve van	22 132	3255	1618	0,0731
V-(t)atik/-(t)etik	16 940	99	61	0,0036
V-va/-ve lesz	3741	1286	707	0,1889



2. ábra. Globális produktivitás a beszélt nyelvi korpuszban

- A legmagasabb globális produktivitás a kézi vizsgálatban is a *V-va/-ve van* sémánál található.
- A legmagasabb potenciális produktivitása a *V-va/-ve lesz* konstrukciónak van, csakúgy, mint az automatikus elemzés eredményeinél.
- *V-(t)atik/-(t)etik* globális produktivitása a legalacsonyabb, ebben a vizsgálatban is mindkét tengelyen rendkívül alacsony (a többinél a típusgyakoriság tekintetében nagyságrendekkel alacsonyabb, a potenciális produktivitás tekintetében a *V-ódik/-ődik* sémát kivéve szintén nagyságrendekkel alacsonyabb) értékek kapcsolódnak ehhez a sémához.
- *V-ódik/-ődik* típusgyakorisága magasabb, mint a *V-va/-ve lesz* sémáé, de potenciális produktivitása jóval alacsonyabb; a hapaok között gyakori a passzív jelentésű konstrukció (vö. 5.). A 4.1. részben említetteknek megfelelően a *V-ódik/-ődik* sémára vonatkozó eredményeket azonban fenntartásokkal kell kezelni az automatikus feldolgozás miatt.

További, elsősorban nem kvantitatív jellegű eredmények:

- a *V-va/-ve van* konstrukció nem csak tranzitív és a generatív szakirodalomban „unakkuzatívnak” nevezett (például *el van sápadva, ki van fáradva*) igei bemenettel fordul elő (vö. Alberti 1996, Kenesei 1998), hanem ún. „unergatív”<sup>26</sup> (például *salakoz/va van*) igékkel is (vö. Németh 2012 is), lásd a (17) példát.

(17) Amikor volt a nagy havazás, akkor nem tudtam jönni. Egyetlen alkalommal, mert nagy volt a hó. *És nem volt salakozva*, felolvadt, és nem fogtak a láncok se, úgyhogy nem tudtam jönni. (doc#2608)

- sőt nem csak telikus igékkel, hanem **atelikus** igékkel is, például

(18) éreztetve/sulykolva/motiválva/lebegtetve/népszerűsítve/foglalkoztatva van,

- valamint nem csak rezultatív/perfektív, hanem **imperfektív** jelentésű konstrukciókban is előfordul, lásd a (18)–(19) példákat.

(19) folyamatosan szorongatva vagyunk

<sup>26</sup> Az „unergatív” és „unakkuzatív” igék elhatárolási kérdéseihöz vö. Halm (2012).

Az imperfektív jelentés részben, de nem kizárólag, a **V-va/-ve van szórendű** variációhoz köthető. A *V-va/-ve van* szórendű konstrukciók a konstrukciós séma megvalósulásainak nem elhanyagolható részét: a tokengyakoriság mintegy 10%-át, a típusgyakoriság mintegy 15%-át teszik ki.

## 6. A hapaxok és a nagy gyakoriságú típusok kvalitatív vizsgálata

### 6.1. A hapaxok szerepe a (potenciális) produktivitásban

Térjünk vissza egy kicsit a választott elméleti keretnek és a vizsgálati módszernek a bevezetőben már taglalt kapcsolatához. Pusztán matematikailag nézve a potenciális produktivitási mutatóban a hapaxok száma a számlálóban található, azaz egyenes arányosságban áll a potenciális produktivitás fokával, míg a nevezőben álló tokengyakoriság fordított arányosságban áll a produktivitási mutatóval, tehát matematikailag csökkenti azt. Ez Baayen (1992) szerint összefüggésbe hozható a pszicholingvisztikai kutatások szófeldolgozáshoz kapcsolódó eredményeivel: a ritka komplex szavak feldolgozása lassabb, mivel feltehetőleg a szabály aktiválásával férünk hozzá a szó jelentéséhez, míg a gyakoribb komplex szavak feldolgozása gyorsabb, mivel ezek erős önálló lexikai reprezentációval rendelkeznek. Bybee (1985; 1988) modellje is hasonlóképpen erős lexikális autonómiát tulajdonít a gyakori, míg gyenge autonómiát a ritka komplex szavaknak, utóbbiak esetében azonban nem egy szabály működtetését, hanem az elmében tárolt hasonló hangzású és jelentésű példányok által alkotott séma aktivációját feltételezi.

A nagyobb fokú lexikális autonómiával rendelkező elemek esetében lexikalizáció figyelhető meg, ami önálló reprezentáció létrejöttével és gyakran a séma/szabály alapján levezethető („kompozicionális”) jelentés megváltozásával jár. Baayen (1992: 140) szerint azonban vannak szabályos jelentésű nagy tokengyakoriságú elemek, és ezek erősítik a morfológiai szabály produktivitását. A kérdés a mi esetünkben az, hogy mi ez a „szabály”, mi a sémához kapcsolható jelentés, illetve mely jelentés(ek)ben produktív egy-egy vizsgált séma. Mind Baayen, mind Bybee érvelése alapján (vö. Kalivoda 2021 is) feltételezhető, hogy az alacsony gyakoriságú elemek, jelen esetben a hapaxok vizsgálatával információkat szerezhetünk a szerkezet produktivitását biztosító elemekről, valamint a sémához kapcsolódó, szabályosságként felfogható műveletről. Ettől a feltételezéstől eltér Kay (2013) nézőpontja, aki szerint a hapax jellegű adatok eltérnek a normától, ezért nem játszanak szerepet

a konstrukcionálódásban (Simon 2024). A jelen tanulmányban – Schultink (1961) Baayen–Lieber (1991) által alkalmazott definíciójának megfelelően – a nagy számban, nem szándékosan létrehozott új lexémák jelentőségét emeljük ki. Kognitív keretben a „szándékolatlanság” kritériumát leginkább alacsony metapragmatikai tudatossággal tudnánk jellemezni, melyet azonban a korpuszelemzés keretében nem tudunk elemezni.

A fenti kettősség rámutat a „hapax legomenon” terminus poliszémiájára: a filológiai és korpusznyelvészeti definíció jelentősen különbözik egymástól. Míg a filológiai értelemben vett hapax valóban csak egy-egy előfordulással bír akár egy adott nyelvben, nyelvváltozatban (sokszor holt nyelvekről, így lezárt korpuszokról lévén szó), a korpusznyelvészeti értelemben vett hapax egy korpuszban fordul elő egyszer, ahol a korpusz a(z élő) nyelv alapsokaságát hivatott reprezentálni.<sup>27</sup> Természetesen lehetséges, hogy a korpuszban előforduló hapax valóban egy egyszeri előfordulás még a nyelv alapsokaságát tekintve is (azaz csak egyvalaki csak egyszer, az adott mintában használta az adott kifejezést), de ennek – a korpusz nagyságától függően – elég csekély a valószínűsége.

A korpusz tehát csak modellál, az, hogy benne egy konstrukció hapax legomenon vagy más igen alacsony gyakoriságú előfordulás, sokszor csak véletlen kérdése. Baayen (2009: 912) mindenesetre megállapítja, hogy a hapaxok esetén ritka szavakról van szó, amelyek lexikális reprezentációja gyenge, ezért a legvalószínűbb, hogy e szavak megértése és produkciója során jobban támaszkodunk a szabályalapú feldolgozásra, míg a magas gyakoriságú szavak esetében erős a lexikális reprezentáció, ezeknél csökken a szabályalapú feldolgozás jelentősége. A használatalapú elméleti keretben természetesen inkább konstrukciós sémáról, mintsem szabályról beszélhetünk, Bybee elméletében „a morfológiai szabályok [...] nem szimbolikus szabályként, hanem lexikai kapcsolatokból kirajzolódó sémaként értendők” (Tanos 2008). Amikor „szabályalapú” feldolgozásról beszélünk, elsősorban a szabályszerű (sémát aktiváló) feldolgozásra helyezük a hangsúlyt az autonóm reprezentáció aktiválásával szemben.

A szabályosság tekintetében szerepet játszik még a bázisnak és a derivátumnak az egymáshoz viszonyított gyakorisága (lásd Baayen 2009): minél nagyobb a bázis gyakorisága a derivátuméhoz képest, annál fontosabb

---

<sup>27</sup> Zeldes (2012: 60, 39. l.) szerint a hapax legomenon kifejezést a klasszikus görög művekkel kapcsolatban a szokatlan, homályos jelentésű, sokszor archaikus szavakra vonatkoztatva használták, míg a modern korpuszokban a legtöbb neologizmus hapax lehet.

szerepet játszanak a szabályok, és fordítva: minél gyakoribb a derivátum a bázishoz képest, annál nagyobb az esélye az önálló reprezentáció létrejöttének. Ezt a hatást kollostrukciós vizsgálattal lehetne jól megfigyelni (vö. 8. pont).

## 6.2. A hapaxok kézi elemzése igenem szerint

Az alábbiakban egy, a gépi elemzés (lásd 4.) során nyert hapaxokból véletlenszerűen kiválasztott, 4x200 elemszámú (minden sémához 200 elemet tartalmazó) minta elemzését mutatjuk be. A minta elemzését kézzel végeztük Microsoft Excel táblázatokban, a kontextus figyelembevételével. Az elemzés során kizártuk a helyesírási vagy annotációs anomália miatt keletkezett hibás találatokat (ez a V-(t)atik/-(t)etik séma esetében kiugróan magas volt), például elgépelés miatt lett hapaxként besorolva a (20) mondatban a *kisodródik* („kisodpródott”) szó, és hibás annotálás miatt tekintette a program a V-va/-ve lesz konstrukcióhoz tartozónak a *pihenve – lesz* kifejezést (20):

- (20) A mexikói *kidodpródott* a Sauberrel, de vissza tudott térni a pályára és a boksza. (doc#2333)
- (21) Mígnem a szülei házában *pihenve* figyelmes lett egy érdekes hírre. (doc#2338)

A találatok kategorizálásánál a következő csoportosítást alkalmaztuk: főként a V-ódik/-ődik sémánál jellemzően mediopasszívnak tekintettük a nem szándékos ágensi, de ráhatott jegyekkel rendelkező használatokat (például *kirotálódhatott*, *elfuserálódik*), passzívnak az ágens által végrehajtott, átperspektivált folyamatokat (például *idevarázsolódik*, [*kilincs*] *lenyitódik*), mediálisnak az alany szándékától független állapotot, folyamatot (például *stabilizálódik*, *elburjánozódik*) és aktívnek a saját erőforrásból végrehajtott cselekvéseket (például *hátraérdeklődik*, *kászálódik*). Az alábbi táblázatban a szortírozás utáni abszolút elemszámok, valamint százalékos értékek láthatók (4. táblázat).

4. táblázat. Hapaxok konstrukciós igeneme az egyes képzési sémákban, abszolút elemszám és százalékos megoszlás

	V-(t)atik/-(t)etik		V-va/-ve lesz		V-va/-ve van		V-ódik/-ődik	
<b>passzív</b>	55	95%	128	88%	96	75%	43	35%
<b>medio-passzív</b>	0	0%	0	0%	1	1%	17	14%
<b>mediális</b>	3	5%	18	12%	31	24%	58	47%
<b>aktív</b>	0	0%	0	0%	0	0%	6	5%
<b>összes elemszám</b>	58	100%	146	100%	128	100%	124	100%

A legnagyobb százalékban passzív jelentésben a V-(t)atik/-(t)etik séma fordult elő (az összes előfordulás 95%-a passzív jelentésű); ezzel a sémával mediális jelentésű hapaxot csak a *születik* és *méltóztatik* ige összetételeiben, valamint egy nyilvánvalóan archaikus előfordulásban (*megszabadultatik*) találunk. A passzív jelentésű hapaxok százalékos arányát tekintve második helyen a V-va/-ve lesz konstrukció áll (88% passzív jelentésű használattal). A V-va/-ve van séma esetében már a hapaxok mintegy negyedénél mediális vagy mediopasszív használatot regisztráltunk, háromnegyede passzív jelentésű. A V-ódik/-ődik séma eredményeiben a passzív és mediopasszív jelentésű elemek számának összege meghaladja a mediális elemek számát. A kisszámú aktív megvalósulás szándékos, önerőből történő cselekvésre utal (például *bekászálódik*).

### 6.3. A magas gyakoriságú elemek

A 2. részben említésre került, hogy feltehetőleg nemcsak az alacsony, de a magas gyakoriságú elemek sémajelentésének is kitüntetett szerepe van a sémák produktivitásában. Bár ezen elemeket inkább egységként tároljuk és dolgozzuk fel, Baayen (2009) megállapítása szerint a magas gyakoriságú, rendszeres képzésű (transzparens) elemek jelenléte növeli az adott konstrukciós séma produktivitását.

Jelen tanulmányban a magas gyakoriságú elemek beható vizsgálatára nem került sor, ennek egyik oka az, hogy a konstrukciós jelentések igen erősen kontextusfüggőek. A gépi vizsgálat során nyert, konstrukciós sémánként tíz-tíz leggyakoribb lemma alapján a következőket lehet megállapítani:

A V-(t)atik/-(t)etik sémában használt 10 leggyakoribb lemmából 7 nem passzív, hanem mediális jelentésű (például *születik*, *viseltetik* 'valamilyen

érzülettel van, *méltóztatik*), ez éles kontrasztot mutat a hapaxok által kirajzolt rendszeres sémajelentéssel, mely 95%-ban passzív. Ez a jelenség arra utal, hogy a Baayen által említett tendencia – miszerint a magas gyakoriságú, szabályos képzésű elemek erősítik a séma produktivitását (vö. Baayen 1992: 140, lásd 2.) – „megfordítása” is igaz lehet, tehát látható, hogy a leggyakoribb elemek nem transzparenssek, és ezzel párhuzamosan megállapítható a konstrukció alacsony globális produktivitása.

A *V-ódik/-ódik* séma esetében többnyire cselekvő igéből képzett, de már jelentésváltozáson átment konstrukciókat figyelhetünk meg, például *folytatódik*, *megkezdődik*, *befejeződik*; már nem transzparens a 9. és 10. leggyakoribb lemma, a *fejlődik* és az *aggódik*.

A *V-va/-ve van* sémával használt 10 leggyakoribb lemma között vannak mediális jelentésűek is, például *meg van győződve*, *fel van készülve*. Ahogy ezt Modrián-Horváth (2024) is kifejti, például a *meg van győződve* konstrukció esetében is megfigyelhető a *meggyőződik* (*valamiről*) ige jelentésétől való nagyfokú eltérés. Kontextusfüggő igenemet tételezhetünk fel az *el van foglalta* konstrukciónál (például a *vár el van foglalta* passzív jelentésű, míg a *valaki el van foglalta* külső hatásra létrejövő belső erődinamikájú, ezért inkább mediális jellegű „elfoglaltságra” utal, vö. *elfoglalja magát*), és 5 passzív jelentésű konstrukció is található a leggyakoribb lexémacsoportban. Mivel e csoportban a hapaxok vizsgálata is jelentős arányban (25%) mutatott mediális konstrukciós jelentést, feltételezhető, hogy a *V-va/-ve van* séma kevésbé specifikus a passzív jelentésre nézve, mint a *V-(t)atik/-(t)etik*, illetve a *V-va/-ve lesz* konstrukciók.

A *V-va/-ve lesz* sémával létrehozott leggyakoribb konstrukciók között kizárólag tranzitív kiinduló igéből képzett konstrukciókat találunk, melyek többnyire passzív jelentésűek (például *biztosítva lesz*, *ki lesz alakítva*); a *meg lesz lepve* kontextusfüggően mediális jelentésű is lehet. Látható, hogy a konstrukció passzív jelentése, amely a hapaxok vizsgálatánál is a második legerősebbnek bizonyult (88%) itt a nagy gyakoriságú elemek passzív jelentésével társul, amely összességében a legerősebb passzív jelentésű produktív sémát adja ki.



## 7. Az eredmények diszkussziója

### 7.1. Az egyes sémák globális produktivitása

Az eredmények értelmezése során némileg problémássá válik az, hogy a globális produktivitás egy kétdimenziós, „vizuális természetű” mérési módszer, amely a kapott számok természete alapján (a potenciális produktivitás 0 és 1 közötti szám, míg a realizált gyakoriság sokszor többes nagyságrendű) nem redukálható egyetlen értékre. Mi alapján lehet eldönteni, hogy két konstrukció közül melyik a produktívabb? Baayen (1993: 190) arra a megállapításra jut, hogy a magasabb típusgyakoriság előnyt élvez, mert ez egy elsődleges, közvetlenül a korpuszadatból származó érték, míg a potenciális produktivitás egy származtatott érték, amelyet a tokengyakoriságok továbbelemzése útján kapunk. Tehát ha mindenáron egy globálisan „legproduktívabb konstrukciót” akarunk megnevezni, a *V-va/-ve van* szerkezetet van okunk annak tekinteni. Azonban a potenciális produktivitási érték az, amelynek inkább van prediktív értéke, mivel a magas típus-token arány önmagában akár egy korábban produktív séma „lenyomata” is lehet a korpuszban (vö. Stefanowitsch 2020), és ugyanezt feltételezhetjük a magas típusgyakorisággal kapcsolatban is. A potenciális produktivitás szempontjából megfigyelt értékek és arányok azt mutatják, hogy a *V-ódik/-ődik* és a *V-(t)atik/-(t)etik* konstrukciók igen alacsony potenciális produktivitással rendelkeznek; ezen értékek nem egészen kétszerese, de így is 0,01 alatti a *V-ni fog* potenciális produktivitási mutatója. A kontrollkonstrukcióhoz képest majdnem háromszor produktívabb a *V-va/-ve van*, és ennél is majdnem két és félszer produktívabb a *V-va/-ve lesz* a potenciális produktivitás szempontjából az egész korpuszra nézve. A *V-ódik/-ődik* és a *V-ni fog* értékei talán meglepőek lehetnek, tekintettel arra, hogy az előbbit a *V-(t)atik/-(t)etik* egyre inkább terjedő alternatívájának tekinti a szakirodalom egy része, illetve a *V-ni fog* szerkezettel kapcsolatban inflexiós jellegénél fogva a passzív konstrukciónál magasabb globális produktivitást várhatnánk. A potenciális produktivitás az alkalmazott képlet alapján fordítottan arányos a tokengyakorisággal, és minél nagyobb a vizsgált korpusz, annál alacsonyabb értéket mutat (az összefüggést problematizálja Gaeta 2007 is). Nem zárhatjuk ki a lehetőségét annak, hogy ezt a tokenszenzitivitást figyelhetjük meg az egyes konstrukciók értékét illetően is, ezt cáfolni látszik azonban az a tény, hogy a *V-(t)atik/-(t)etik* és a *V-va/-ve van* tokengyakorisága között nincs nagyságrendi, sem többszörös különbség (mintegy 15%-os az eltérés),

a potenciális produktívásban mégis drámai eltérés mutatkozik. A *V-ódik/-ődik* séma mediális, mediopasszív és passzív jelentéseket fog át, melyek nem különülnek el egymástól élesen. Magas típusgyakorisága összefüggésbe hozható e széles jelentésbeli spektrummal, valamint egy korábbi magasabb fokú produktivitással. A potenciális produktivitást illetően a hapaxok vizsgálata alapján valószínűsíthető, hogy a konstrukciós jelentésnek csak bizonyos (passzív és mediopasszív) szegmensei mutatnak nagyobb fokú produktivitást, míg a kevésbé produktív, de igen magas példánygyakoriságú mediális konstrukciók csökkentik a potenciális produktívási mutatót.

Az egész korpuszra vetítve tehát a legmagasabb globális produktívás a *V-va/-ve van* sémánál található, a konstrukcióval kapcsolatos közismert nyelvi babonák ellenére. A *V-va/-ve lesz* típusgyakorisága ezzel szemben az egész korpuszban alacsonyabb, melynek számos oka lehet, például a személytelen T/3 szerkezet (lásd F. Gulyás 2016) használata a dinamikus passzív helyett, a rugalmas szórend kiaknázása a páciens előtérbe helyezésére dinamikus események konstruálásánál. A *V-va/-ve lesz* séma megvalósulásainak egy része szemantikai kapcsolódások révén részben a *V-va/-ve van* konstrukcióhoz tartozhat (a *V-va/-ve lesz* konstrukcióból vett hapaxmin tában mintegy 53% volt a biztosan a *V-va/-ve lesz*-hez tartozó, múlt idejű, kijelentő módú konstrukciók aránya). Ha a két konstrukciót egyesítenénk, a *V-va/-ve van* magas példánygyakorisága miatt a *V-va/-ve van* értékeihez közeli, kissé magasabb potenciális produktívású és alacsonyabb típusgyakoriságú értékeket kapnánk. Az informális regisztert képviselő személyes alkorpuszban azonban a *V-va/-ve lesz* konstrukciónak nemcsak a potenciális, de a típusgyakorisága is kiugróan magas (a típusgyakoriság meghaladja a *V-ódik/-ődik* sémáét). Bár az MNSZ2 alkorpuszainak elnevezése azt engedné sejtetni, hogy a beszélt nyelvi alkorpusz szövegei állnak a spontán mindennapi társalgás nyelvhasználatához a legközelebb, ez az alkorpusz (ahogy az 5. fejezetben említettük) rádiós beszélgetések leiratát tartalmazza, melyek csak részben spontán megnyilatkozások. Ezért joggal feltételezhetjük, hogy a személyes alkorpusz adatai az ún. beszélt nyelvihez hasonlóan, sőt esetleg jobban tükrözik a hétköznapi nyelvhasználatot, tehát a *V-va/-ve lesz* sémának kiemelt szerepe van a spontán, informális nyelvhasználatban a (dinamikus) passzív eseménykonstruálásban.

A *V-(t)atik/-(t)etik* séma produktívásának megítélése a szakirodalomban nem volt egységes, ennek oka abban keresendő, hogy az említett grammatikák más-más produktívási fogalommal dolgoznak, így a potenciális szavak vagy a filológiai értelemben vett hapaxok létrehozásának lehetősége

a kompetencianyelvtanok számára bizonyítékként szolgálhat egy adott képző vagy séma produktivitására. A *V-(t)atik/-(t)etik* produktivitásának kvantitatív jellegű mérése során megerősítést nyert, hogy a képzési séma produktivitása igen korlátozott, ez egybevág a grammatikák többségében megfogalmazott feltételezésekkel (vö. Kenesei et al. 1998, Ladányi 2017) és a tanulmány szerzőinek intuíciójával is. E ponton kiemelnénk, hogy a nyelvi intuíciónak való megfelelés Baayen–Lieber (1991) megközelítésében is fontos tényező. Ugyanakkor meg kell említeni azt aényt is, hogy a *V-(t)atik/-(t)etik* séma alacsony típusgyakorisága és potenciális produktivitása ellenére közepes tokengyakorisággal bír. A tokengyakoriság hatása akár erősíthetné is a produktivitást, amennyiben a séma elsődleges, passzív jelentésmintázatára illeszkedne (melyet a hapaxok vizsgálata is kimutatott), azonban a legnagyobb példányszámú típusok (például *születik*) jellemzően nem passzív jelentésűek.

A produktivitás kutatásában jelentős tényezőként merül fel a nyelv szociokulturális komplexitásából adódó, regiszterek közti különbség vizsgálata, melyet az MNSZ2, illetve az MNSZ2-UD egyes alkorpuszai reprezentáltak. Megfigyelhettük, hogy a *V-va/-ve lesz* konstrukció produktivitása a teljes korpuszon mért értékeknél jóval alacsonyabb a szépirodalmi, és sokkal magasabb a személyes alkorpuszban. A tudományos alkorpuszban a *V-va/-ve van* konstrukció bizonyult feltűnően produktívnek mindkét dimenzió mentén. Az összehasonlítási alapként használt *V-ni fog* konstrukció a hivatalos, a sajtónyelvi, valamint a (rádiós szövegeket tartalmazó) ún. beszélt nyelvi alkorpuszokban kimagasló típusgyakorisággal és a beszélt nyelvi alkorpuszban kimagasló tokengyakorisággal fordult elő, míg a teljes korpusznál alacsonyabb volt a típusgyakoriság a tudományos alkorpuszban.

## 7.2. Szintetikus és analitikus konstrukciók globális produktivitása

A vizsgálatok megkezdése előtt felmerült az a kérdés, hogy összehasonlítható-e a szintetikus és analitikus konstrukciók produktivitási indexe. Bár a szakirodalomban nem ismeretlen az eredetileg szintetikus kifejezőeszközök produktivitásának mérésére kifejlesztett globális produktivitási index szintaktikai jellegű konstrukciókra való használata, a két konstrukciótípus mérőszámainak egymással való összevetése mégis metodológiai kihívást, újítást jelentett. Vizsgálatunkba ezért bevontunk összehasonlítási alapként egy nem passzív jelentésű, grammatikalizálódott, igeidőjelölő analitikus konstrukciót, a *V-ni fog* konstrukciót. Az eredmények azt mutatták, hogy a *V-ni fog* típusgyakorisága kissé magasabb a *V-va/-ve van* konstrukciónál

(a *V-ódik/-ődik* és a *V-va/-ve van* között van), potenciális produktivitása viszont jóval alacsonyabb mindkét vizsgált passzív analitikus sémánál. Ebből arra lehet következtetni, hogy a konstrukció szintetikus vagy analitikus jellege nem áll közvetlen összefüggésben sem a realizált, sem a potenciális produktivitás mértékével, ami megnyugtató a jelen vizsgálat metodológiájával kapcsolatban. A *V-ni fog* séma értékeivel való összehasonlítás eredményei újabb kérdéseket vetnek fel a *V-va/-ve van* konstrukció státuszára nézve: ez a konstrukció is az igei paradigma részét képezi, a morfológiai leírásokban van a helye?

### 7.3. Kézi és gépi elemzések eredményei

A 4. fejezetben gépi elemzéssel nyert, a beszélt nyelvi alkorpuszra vonatkozó adatokat összevetettük az 5. fejezetben található kézi elemzéssel nyert adatokkal. Ennek során azt találtuk, hogy a sémák egymáshoz viszonyított produktivitása nem változott, ellenben az abszolút értékekben jelentős különbségek is mutatkoztak. A kézi elemzésnél problémát jelentett, hogy a kiemelkedően magas tokengyakoriságú *V-ódik/-ődik* sémánál a jelen projekt keretében kivitelezhetetlen volt az adatok kézi feldolgozása, így az e sémára vonatkozó adatok kizárólag gépi elemzésen alapulnak. – Az összehasonlítás eredménye a teljes vizsgálat validitására nézve biztató az egyes konstrukciók egymáshoz viszonyított produktivitásával kapcsolatban.

A nagy adatmennyiség miatt a teljes korpuszra, illetve a hat alkorpuszból ötre nézve kizárólag a gépi elemzéssel lehetett adatokat nyerni. Mindamellett néhány jelenség vizsgálata kézi elemzést igényelt, ilyen (az eredmények validálásán túl) a hapaxok jelentéseinek vizsgálata. Ezenkívül a kézi korpuszvizsgálat során derült fény arra, hogy a *V-va/-ve van* szórendű konstrukciók (például *lebegtetve van*), melyek egy részét az automatikus elemzésnél (4. fejezet) a magas hamis pozitív találati arány miatt ki kellett zárni a vizsgálatból, a konstrukciós séma jelentős részét: a tokengyakoriság mintegy 10%-át, a típusgyakoriság mintegy 15%-át teszik ki (lásd 5.2). A *V-va/-ve van* szórendű konstrukciók feltehetőleg külön típust képviselnek aspektuális tekintetben a *V-va/-ve van* konstrukciókon belül, erre szintén a kézi elemzés során lettünk figyelmesek. A manuális korrekció során ismertük fel azt is, hogy az igekötős *V-va/-ve van/lesz* szórendű találatok szinte kivétel nélkül hamis pozitívak. Mindez alátámasztja annak fontosságát, hogy ha a vizsgálat nem is folytatható le egészében kézi vizsgálattal, mégis fontos az adatok beható ismerete.

#### 7.4. Hapaxok és nagy gyakoriságú elemek vizsgálata

A hapaxok és nagy gyakoriságú elemek sémajelentésének vizsgálata hasznosnak bizonyult annak megállapításánál, hogy mely jelentésekben a legproduktívabbak a vizsgált sémák. A hapaxokból nyert minta, valamint a leggyakoribb igelemmák vizsgálata alapján megállapítható, hogy a *V-(t)atik/-(t)etik* produktivitása (mely igen alacsony) szinte kizárólag a passzív jelentésű elemekhez köthető, míg a legnagyobb példányszámú lexémák mediális jelentésűek. A másik legstabilabban passzív jelentésű séma a *V-va/-ve lesz* (a mintában 88%-ban passzív jelentésű hapaxokkal és többnyire passzív jelentéssel a leggyakoribb lemmák között), ezt követi a *V-va/-ve van* (itt 75%-ban volt passzív a hapaxok jelentése, és a leggyakoribb lemmák között is több mediális bemeneti igét találtunk). A *V-va/-ve van* és *V-va/-ve lesz* jelentős különbségeket mutat passzív/mediális megoszlás szempontjából, miközben tudvalevő, hogy a *lesz* tövű szavak gyakran a *van* szuppletív alakjaiként szerepelnek (*lenne, lesz, lett volna, legyen* stb.).

A 2. fejezetben írottak alapján a *V-va/-ve van*, illetve *V-va/-ve lesz* konstrukciók egyértelmű elkülönítése nehézségekre ütközik mind kézi, mind gépi vizsgálat során: az utóbbi esetben a nagy adatmennyiség is közrejátszik abban, hogy csak nagyon nehezen, vagy egyáltalán nem kivitelezhető a tövek elkülönítése, a kézi elemzésnél pedig a kontextus alapján számos esetben nem egyértelmű, hogy a *van* vagy a *lesz* ige paradigmájához van közelebb az aktuális használat. Az egyetlen egyértelmű esetet a *V-va/-ve lesz* kijelentő módú, múlt idejű alakjai képviselik. A hapaxok kézi vizsgálata alapján ezek aránya mintegy 53% a mintában, azaz a *lesz* lemmához sorolt előfordulások több mint feléről van szó (76 eset a 146-ból; a múlt idejű használatok aránya hasonló a *V-va/-ve volt* vonatkozásában). Figyelemre méltó módon ezek közül mindössze egy tartozott a mediális jelentésű konstrukciók közé (*fel lettünk világosodva*), az összes többi *V-va/-ve lett* múlt idejű, kijelentő módú paradigmatis alak passzív jelentéssel bírt. Ez az eredmény abba az irányba mutat, hogy valóban rendszeres különbség lehet a *V-va/-ve van* és *V-va/-ve lesz* konstrukciók sémajelentései között.

A *V-ódik/-ődik* séma, melyben sokan az új szintetikus passzívképző lehetőségét látják, az esetek mintegy 35%-ában mutatott passzív jelentést a hapaxok között, ehhez járul még mintegy 14%-nyi mediopasszív (ráhatott) használat. A produktivitási tendenciák megállapításánál érdemes lehet figyelembe venni, hogy a mediopasszív tekintett (ráhatott) előfordulásokat más nyelvekben tipikusan passzív alakokkal fejezik ki.

## 8. Összegzés és kitekintés

A tanulmány célja négy passzív jelentést mutató konstrukció korpuszalapú vizsgálata volt. A globális produktivitás kvantitatív vizsgálatának eredménye leginkább a hapaxok és a magas gyakoriságú elemek vizsgálatával összevetve értékelhető a vizsgálni kívánt jelenségre nézve. Az elemzés során azt lehetett megállapítani, hogy a legmagasabb globális produktivitás a *V-va/-ve van* sémánál található, mely a hapaxok vizsgálatá alapján túlnyomórészt (a minta 75%-ában) passzív jelentést tartalmaz. A személyes alkorpuszban, amely feltehetőleg jól tükrözi a spontán hétköznapi nyelvhasználatot, a *V-va/-ve lesz* globális produktivitása is kiugróan magas, e séma jelentése a hapaxok mintegy 88%-ában passzív, ezért arra következtethetünk, hogy a passzív jelentést tartalmazó nem modális, dinamikus predikatív megkonstruálására elsősorban a *V-va/-ve lesz* séma alkalmas produktívan. Az egyértelműen ehhez a sémához tartozó *V-va/-ve lett* múlt idejű, kijelentő módú paradigmatis alakoknál csaknem kivétel nélkül passzív sémajelentést találunk, ami erősíti ezt a feltételezést. A vizsgálatokban legalacsonyabb globális produktivitása – a nyelvi intuíciónak megfelelően – a *V-(t)atik/-(t)etik* szinte kizárólag passzív jelentésű sémának volt. A *V-odik/-odik* séma nagy típusgyakorisággal, de igen alacsony potenciális produktivitással rendelkezik, és a hapaxok kvalitatív vizsgálata alapján a séma jelentése megosztott: csaknem 50%-ban mediális, 35%-ban passzív, és mintegy 15%-ban mediopasszív. A magas gyakoriságú elemek vizsgálata a hapaxokon a fent említettekkel egybevágó eredményeket mutatott a sémák jelentésére, illetve a produktív sémajelentésekre vonatkozóan.

A *V-va/-ve van* séma vizsgálata során az eddigi leírásokban feltételezettnél gazdagabb spektrumú igei bázisokra lettünk figyelmesek: kiinduló igeként atelikus igék realizációi is megjelennek, gyakran egy speciális szórendi mintázattal (*motiválva/lebegtetve van*). A *V-va/-ve van* és *V-va/-ve lesz* sémáknál egyaránt megfigyelhettük a személytelen passzív két tulajdonságát (vö. F. Gulyás 2016: 119kk.): az alanytalan passzívumot és az intranszitiv igékből képzett konstrukciókat.

A globális produktivitásra vonatkozó gépi és kézi elemzések összevetése hasonló eredményeket mutatott az egyes sémák egymáshoz képesti produktivitását tekintve, azonban a számszerű adatokra nézve eltérések mutatkoztak. A kutatás kvalitatív elemzésből nyert eredményei is aláhúzzák az adatok beható ismeretének fontosságát. A hapaxok manuális vizsgálatával értékes eredményeket nyertünk az egyes sémák megvalósulásainak igenemével

kapcsolatban, így fény derült például arra, hogy a *lesz* ige múlt idejű tövével alkotott *V-va/-ve lesz* konstrukciók gyakorlatilag mindegyike passzív jelentésű, illetve összefüggéseket találtunk a *V-va/-ve van* konstrukció szórendi megvalósulása és aspektuális tartalma között. Egy másik metodológiai kérdés volt a szintetikus és analitikus konstrukciók globális produktivitási indexének összehasonlíthatósága. Itt az összehasonlítási alapként vizsgált *V-ni fog* elemzése alapján (magas típusszám, alacsony potenciális produktivitás) úgy tűnik, nincs összefüggés a szintetikus/analitikus jelleggel.

A vizsgálat további kérdéseket, lehetőségeket nyit a jövőbeni kutatások számára. Ilyen kérdés például a *van és lesz*, *V-va/-ve van* és *V-va/-ve lesz* paradigmák szemantikai kapcsolatainak vizsgálata, az utóbbi analitikus konstrukciókat a jelen tanulmányban metodológiai célból különválasztottuk. Elméleti jelentőségű az a megállapítás, hogy a *V-va/-ve van* konstrukció potenciális produktivitása jóval magasabb, a típusgyakorisága közel van a kontrollként vizsgált *V-ni fog* összetett igeidő-konstrukció értékeihez, ez ugyanis felveti azt a kérdést, vajon a *V-va/-ve van* konstrukció is analitikus (összetett, segédigés) igealakként kezelendő, az igei paradigma részét képezi-e. Ennek megítélésénél azt is érdemes tekintetbe venni, hogy egy jelölt igenem- és aspektuális komponensekkel bíró konstrukcióról van szó, amelyvel kapcsolatban nem várnánk, hogy produktivitása felveszi a versenyt egy igeidő-jelölő konstrukcióéval, vagy akár felül is múlja azt.

Nyitott kérdésként kezelhető a passzív és mediális elhatárolásának problémája, amelynél a jelenleg használt kritériumok finomítására lehet szükség, tekintettel a ráhatottságot kifejező mediopasszív használatokra. Itt elsősorban az ágéntív passzív metaforikus és metonimikus kiterjesztése által létrejövő polyszémia behatóbb vizsgálatára lenne szükség az olyan, nem nyíltan hozzáférhető folyamatok konceptualizációjában, amelyekben a folyamat kivitelezője nem állapítható meg (például *letöltődik*, *át lesz irányítva*, *elsáncolódik*). Végül, de nem utolsó sorban, az analitikus konstrukciókkal végzendő kollostrukciós vizsgálat alkalmas lehetne a konstrukció és a benne megjelenő lemmák közötti affinitások kimutatására.

## Irodalom

- Alberti Gábor 1992. Passzivizálási művelet a magyarban. *Néprajz és Nyelvtudomány* 37: 7–45.
- Aronoff, Mark 1976. *Word formation in Generative Grammar*. Cambridge, MA: MIT Press.

- Baayen, Harald 1992. Quantitative aspects of morphological productivity. In: Booij, Geert Evert et al. (eds.): *Yearbook of Morphology 1991*. 109–149. [https://doi.org/10.1007/978-94-011-2516-1\\_8](https://doi.org/10.1007/978-94-011-2516-1_8)
- Baayen, Harald R. 1993. On frequency, transparency and productivity. In: Booij, Geert – van Marle, Jaap (eds.): *Yearbook of Morphology 1992*. Dordrecht: Kluwer Academic. 181–208. [https://doi.org/10.1007/978-94-017-3710-4\\_7](https://doi.org/10.1007/978-94-017-3710-4_7)
- Baayen, Harald 2009. Corpus linguistics in morphology: Morphological productivity. In: Lüdeling, Anke – Kytö, Merja (eds.): *Corpus Linguistics. An international handbook. Vol. 2.* (HSK 29.2). Berlin, New York: Mouton de Gruyter. 899–919. <https://doi.org/10.1515/9783110213881.2.899>
- Baayen, Harald – Lieber, Rochelle 1991. Productivity and English derivation: A corpus based study. *Linguistics* 29: 801–843. <https://doi.org/10.1515/ling.1991.29.5.801>
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins. <https://doi.org/10.1075/tsl.9>
- Bybee, Joan L. 1988. Morphology as lexical organization. In: Hammond, Michael – Noonan, Michael (eds.): *Theoretical morphology: Approaches in modern linguistics*. London: Academic Press. 119–141. [https://doi.org/10.1163/9789004454101\\_010](https://doi.org/10.1163/9789004454101_010)
- Bybee, Joan 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750526>
- Croft, William 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.
- Croft, William 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198299554.001.0001>
- de Marneffe, Marie-Catherine – Manning, Christopher D. – Nivre, Joakim – Zeman, Daniel 2021. Universal Dependencies. *Computational Linguistics* 47 (2): 255–308. [https://doi.org/10.1162/coli\\_a\\_00402](https://doi.org/10.1162/coli_a_00402)
- Diessel, Holger 2014. Usage-based linguistics. In: Mark Aronoff (ed.): *Oxford Research Encyclopedia of Linguistics*. New York: Oxford University Press. <https://doi.org/10.1093/obo/9780199772810-0068>
- Dressler, Wolfgang – Ladányi, Mária 2000. Productivity in word formation (WF): A morphological approach. *Acta Linguistica Hungarica* 47 (1): 103–44. <https://doi.org/10.1023/A:1014010530824>
- E. Abaffy Erzsébet 1978. A mediális igéről. *Magyar Nyelv* 74: 280–293.



- F. Gulyás Nikolett 2016. *Személytelen szerkezetek finnugor nyelvekben funkcionális és tipológiai megközelítés*. Budapest: ELTE.
- Gaeta, Livio 2007. On the double nature of productivity in inflectional morphology. *Morphology* 17 (2): 181–205. <https://doi.org/10.1007/s11525-007-9117-7>
- Goldberg, Adele 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>
- Grétsy László 2014. Megivódik. *Szabad Föld* [2014. 11. 14.] <https://szabadfold.hu/orszag-vilag/2014/11/gretsy-laszlo-megivodik>
- Halm Tamás 2012. Unergatív és/vagy unakkuzatív: a szemelfaktív igék argumentum-szerkezetéről, szemantikájáról és szintaxisáról a magyarban. In: Gécseg, Zsuzsanna (szerk.): *LingDok 11.: Nyelvészdoktoranduszok dolgozatai*. Szeged: JATEPress. 73–88.
- Hilpert, Martin 2013. *Constructional change in English. Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139004206>
- Honti László – H. Varga Márta 2013. *Meg van írva!* A határozói igenév és a létige alkotta szerkezet funkciójáról és háttéréről. In: Honti László (szerk.): *Magyar nyelvtörténeti tanulmányok*. Budapest: Károli Gáspár Református Egyetem, L'Harmattan. 255–266.
- Imrényi András 2017. Az elemi mondat viszonyhálózata. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 664–760.
- Janda, Laura 1990. The radial network of a grammatical category – its genesis and dynamic structure. In *Cognitive Linguistics* 1 (3): 269–288. <https://doi.org/10.1515/cogl.1990.1.3.269>
- Janda, Laura A. 2013. Quantitative methods in Cognitive Linguistics: An introduction. In: Janda, Laura A. (ed.): *Cognitive Linguistics: The quantitative turn – The essential reader*. Berlin, Boston: Walter de Gruyter. 1–32. <https://doi.org/10.1515/9783110335255.1>
- Kalivoda Ágnes 2019. Véges erőforrás végtelen sok igekötős ígére. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*. Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet. 331–344.
- Kalivoda Ágnes 2021. Igekötős szerkezetek a magyarban. Doktori (PhD) értekezés. Budapest: PPKE BTK.

- Kalivoda, Ágnes – Malouf, Robert – Ackerman, Farrell – Sass, Bálint 2024. *Building a dependency treebank from the Hungarian Gigaword Corpus*. Kézirat, megjelenés előtt.
- Kay, Paul 2013. The limits of (Construction) Grammar. In: Hoffmann, Thomas – Trousdale, Graeme (eds.): *The Oxford handbook of Construction Grammar*. New York: Oxford University Press. 32–48. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0003>
- Kádár Edit – Németh Boglárka 2009. Predikatív határozói igeneves szerkezetek csángó beszélt nyelvi szövegekben. In: É. Kiss Katalin – Hegedűs Attila (szerk.): *Nyelvelmélet és dialektológia*. Piliscsaba: Pázmány Péter Katolikus Egyetem. 189–212.
- Kemmer, Suzanne 1993. *The middle voice*. Amsterdam, Philadelphia: John Benjamins. <https://doi.org/10.1075/tsl.23>
- Keszler Borbála (szerk.) 2000. *Magyar grammatika*. Budapest: Akadémiai kiadó.
- Komlósi András 1992. Régensek és vonzatok. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 1. Mondattan*. Budapest: Akadémiai Kiadó. 299–528.
- Kugler, Nóra – Simon, Gábor 2018. The schematization of Hungarian participle-noun compounds. *SKY Journal of Linguistics* 31: 35–69.
- Laczó Krisztina 2022. Adalékok a mediális igés kompozitumszerkezetekhez. In: Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.) *A magyar mondat és kontextuális környezete*. Budapest: ELTE BTK. 76–106.
- Ladányi Mária 2008. Produktivitás és analógia a szóképzésben: a magyar főnévképzés néhány esete. Általános Nyelvészeti Tanulmányok XXII. Tanulmányok a funkcionális nyelvészet köréből. 349–430.
- Ladányi Mária – Tolcsvai Nagy Gábor 2008. Funkcionális nyelvészet. *Általános Nyelvészeti Tanulmányok XXII. Tanulmányok a funkcionális nyelvészet köréből*. 17–58.
- Lakoff, George 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, London: The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226471013.001.0001>
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. I. Theoretical prerequisites*. Stanford California: Stanford University Press.
- Langacker, Ronald W. 2006: Dimensions of defocusing. In: Shibatani, Masayoshi et al. (eds.): *Voice and grammatical relations: In honor of Masayoshi Shibatani*. Amsterdam, Philadelphia: John Benjamins. 115–137. <https://doi.org/10.1075/tsl.65.08lan>

- Lengyel Klára 2000a. A segédigék és származékaik. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó. 252–258.
- Lengyel Klára 2000b. *Az igenevek helye a szófaji rendszerben*. Nyelvtudományi Értekezések 146. Budapest: Akadémiai Kiadó.
- Lewandowska-Tomaszczyk, Barbara 2007. Polysemy, prototypes, and radial categories. In: Geeraerts, Dirk – Cuyckens, Hubert (eds.): *The Oxford handbook of Cognitive Linguistics*. Oxford: Oxford University Press. 139–169.
- Modrián-Horváth Bernadett 2024. Vázlat a magyar passzív jelentésű konstrukciók vizsgálatához a használatalapú nyelvtan szemszögéből. In: Tátrai Szilárd – Laczkó Krisztina – Tolcsvai Nagy Gábor (szerk.): *Magyar igei konstrukciók. Használatalapú függőségi vizsgálatok*. (Megjelenés alatt.)
- Németh Boglárka 2012. *Az aspektus a magyar nyelvben. Különös tekintettel a statikusságra*. Budapest: Tinta Könyvkiadó.
- Novák, Attila – Siklósi, Borbála – Oravecz, Csaba 2016. A new integrated open-source morphological analyzer for Hungarian. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia: European Language Resources Association (ELRA). 1315–1322.
- Oravecz, Csaba – Várad, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík: European Language Resources Association (ELRA). 1719–1723.
- Pethő Gergely – Sass Bálint – Kalivoda Ágnes – Simon László – Lipp Veronika 2022. Igekötő-kapcsolás. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem, Informatikai Intézet. 77–91.
- Rosch, Eleanor 1978. Principles of categorization. In: Rosch, Eleanor – Lloyd, Barbara B. (eds.): *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates. 27–48. <https://doi.org/10.4324/9781032633275-4>
- Schultink, Henk 1961. Produktiviteit als morfologisch Fenomeen. *Forum der Letteren* 2: 110–125.
- Siewierska, Anna 2013. Passive constructions. In: Dryer, Matthew S. – Haspelmath, Martin (eds.) *WALS Online* (v2020.3) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7385533> (<http://wals.info/chapter/107> [2023. 10. 05.] )
- Simon Gábor 2024. Konstruktio és/vagy mintázat: Az azonosítás kérdései egy használatalapú, korpuszvezérelt nyelvleírásban. In: Tátrai Szilárd – Laczkó Krisztina – Tolcsvai Nagy Gábor (szerk.): *Magyar igei konstrukciók. Használatalapú függőségi vizsgálatok*. (Megjelenés alatt.)

- Stefanowitsch, Anatol 2006. Konstruktionsgrammatik und Korpuslinguistik. In: Fischer, Kerstin – Stefanowitsch, Anatol (Hg.): *Konstruktionsgrammatik I: Von der Anwendung zur Theorie*. Tübingen: Stauffenburg.
- Stefanowitsch, Anatol 2020. *Corpus linguistics. A guide to the methodology*. Textbooks in Language Sciences 7. Berlin: Language Science Press. DOI: 10.5281/zenodo.3735822
- Szepesy Gyula 1972. A *lenni* + *-va*, *ve* igeneves szerkezet funkciói. *Magyar Nyelvőr* 96: 404–414.
- Szili Katalin 1999. Valahol a passzívum és a mediálisok között ... (A szenvedő jelentéstartalom változó megjelenésmódja kapcsán). *Magyar Nyelvőr* 123 (3): 350–357.
- Tanos Bálint 2008. Joan Bybee nyelvelmélete. *Általános Nyelvészeti Tanulmányok XXII*: 537–566.
- Tanos Bálint 2012. Határozószó-képzési produktivitás kvantitatív keretben. In: Tolcsvai Nagy Gábor, Tátrai Szilárd (szerk.): *Konstrukció és jelentés. Tanulmányok a magyar nyelv funkcionális kognitív leírására*. Budapest: ELTE. 59–69.
- Tolcsvai Nagy Gábor 2017. Jelentéstan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 207–499.
- Tsötönyi Márton 1831. *Leg hasznosabb és új fel fedezésekkel irott rövid gazdasági munka*. Pest: Wigand Ottó.
- van Marle, Jaap 1985. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris. <https://doi.org/10.1515/9783111558387>
- van Wettère, Niek 2022. The hapax / type ratio. An indicator of minimally required sample size in productivity studies? *International Journal of Corpus Linguistics* 27 (2): 166–190. <https://doi.org/10.1075/ijcl.19114.van>
- Zeldes, Amir 2012. *Productivity in Argument Selection: From Morphology to Syntax*. Berlin, Boston: Walter de Gruyter. <https://doi.org/10.1515/9783110303919>



## Quantitative productivity analysis of Hungarian passive predicative constructions

The aim of the study is to investigate the quantitative productivity of four passive constructions: *V-ódik/-ődik*, *V-(t)atik/-(t)etik*, *V-va/-ve lesz*, and

*V-va/-ve van*. The research is based on automatic and manual analyses using the HGC and the reannotated HGC-UD corpora, and combines the global productivity index (cf. Baayen 1992), which can be calculated in a quantitative framework, with qualitative elements. The Global Productivity Index is a measure often used in usage-based linguistics, which can be calculated using the type and token frequency of constructional schemas and the number of hapaxes legomena fitting into the schemas. The qualitative part of the study was aimed at classifying the sample of hapaxes that best reflect the productive meaning of each schema according to their voice.

The results of the study show that the schema with the highest global productivity in the totality of the corpus is the *V-va/-ve van* construction, the one with the highest type frequency is *V-ódik/-ődik*, while the one with the highest potential productivity is *V-va/-ve lesz*; the results for individual sub-corpora differ partially. The analysis of a random sample of hapaxes shows that the schema with the most persistent passive meaning is the extremely low productivity construction *V-tatik/-tetik*, closely followed by *V-va/-ve lesz*, the *V-va/-ve van* predominantly passive meaning, and finally less than half of the hapaxes have passive meaning in *V-ódik/-ődik*. Further theoretical and methodological considerations are also discussed in the paper.

**Keywords:** Hungarian grammar, voice, passive, corpus linguistics, construction grammar

*Modrián-Horváth Bernadett*

Szegedi Tudományegyetem

bernadett.modrianhorvath@gmail.com

ORCID: 0000-0003-3834-9163

*Kalivoda Ágnes*

HUN-REN Nyelvtudományi Kutatóközpont

kalivoda.agnes@nytud.hun-ren.hu

ORCID: 0000-0003-2520-5523

*Tanos Bálint*

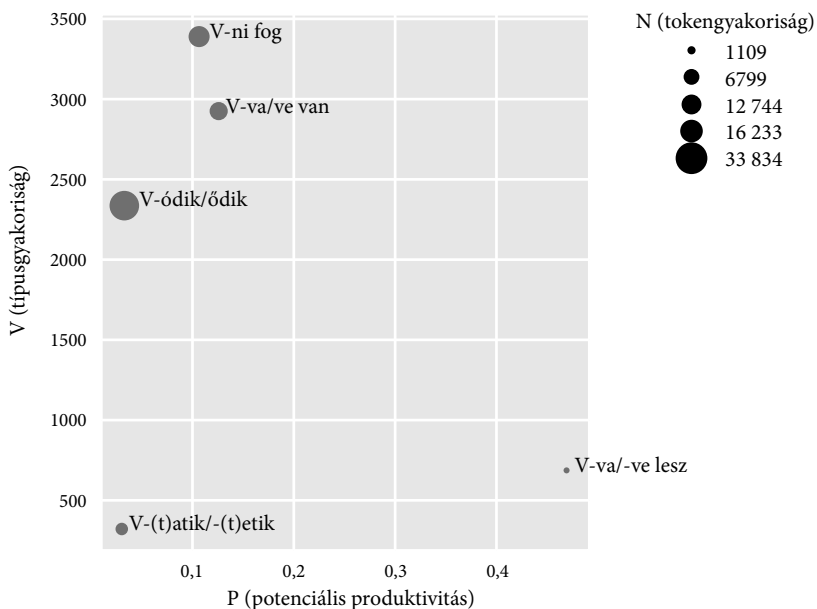
batanos@gmail.com

ORCID: 0009-0005-6070-3490

## I. számú melléklet: Szépirodalmi alkorpusz

5. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n<sub>1</sub>) és potenciális produktivitása (P) a szépirodalmi alkorpuszban

szerkezet	N	V	n <sub>1</sub>	P
V-ódik/-ődik	33 834	2325	1177	0,0348
V-va/-ve van	12 744	2915	1606	0,1260
V-(t)atik/-(t)etik	6799	330	213	0,0313
V-va/-ve lesz	1109	688	518	0,4671
V-ni fog	16 233	3379	1761	0,1085

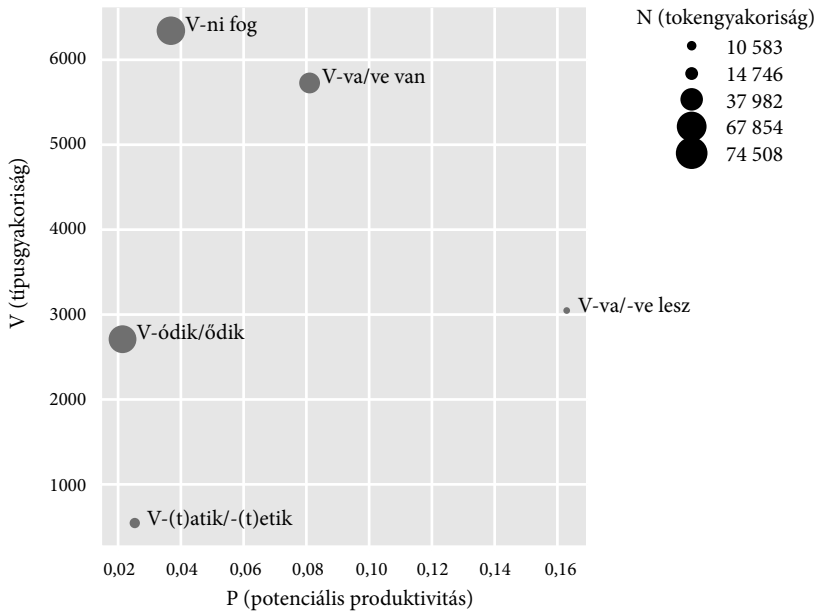


3. ábra. A vizsgált konstrukciók globális produktivitása a szépirodalmi alkorpuszban

## II. számú melléklet: Személyes alkorpusz

6. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma ( $n_1$ ) és potenciális produktivitása (P) a személyes alkorpuszban

szerkezet	N	V	$n_1$	P
V-ódik/-ődik	67 854	2687	1439	0,0212
V-va/-ve van	37 982	5682	3080	0,0811
V-(t)atik/-(t)etik	14 746	535	368	0,0250
V-va/-ve lesz	10 583	3022	1723	0,1628
V-ni fog	74 508	6293	2735	0,0367

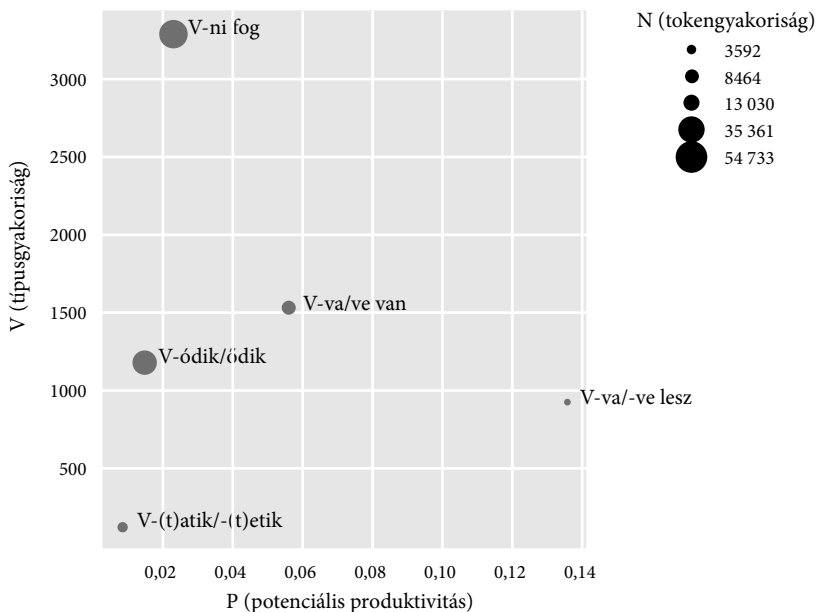


4. ábra. A vizsgált konstrukciók globális produktivitása a személyes alkorpuszban

### III. számú melléklet: Hivatalos alkerpusz

7. táblázat. A vizsgált konstrukciók tokenygyakorisága (N), típusgyakorisága (V), hapaxainak száma (n<sub>1</sub>) és potenciális produktivitása (P) a hivatalos alkerpuszban

szerkezet	N	V	n <sub>1</sub>	P
V-ódik/-ődik	35 361	1179	533	0,0151
V-va/-ve van	13 030	1528	739	0,0567
V-(t)atik/-(t)etik	8464	131	75	0,0089
V-va/-ve lesz	3592	928	489	0,1361
V-ni fog	54 733	3280	1293	0,0236



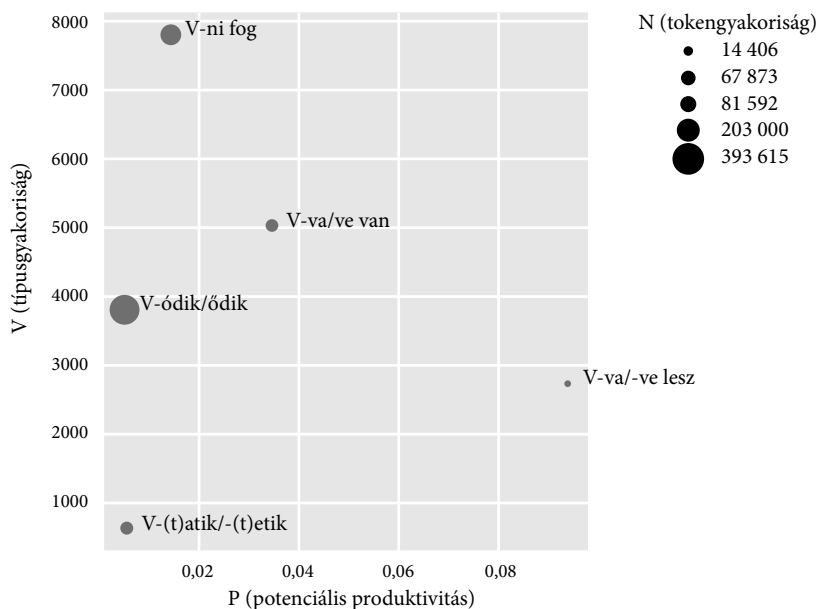
5. ábra. A vizsgált konstrukciók globális produktivitása a hivatalos alkerpuszban



## IV. számú melléklet: Sajtónyelvi alkorpusz

8. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n<sub>1</sub>) és potenciális produktivitása (P) a sajtónyelvi alkorpuszban

szerkezet	N	V	n <sub>1</sub>	P
V-ódik/-ődik	393 615	3768	1800	0,0046
V-va/-ve van	67 873	5008	2336	0,0344
V-(t)atik/-(t)etik	81 592	598	415	0,0051
V-va/-ve lesz	14 406	2696	1354	0,0940
V-ni fog	203 000	7772	2823	0,0139

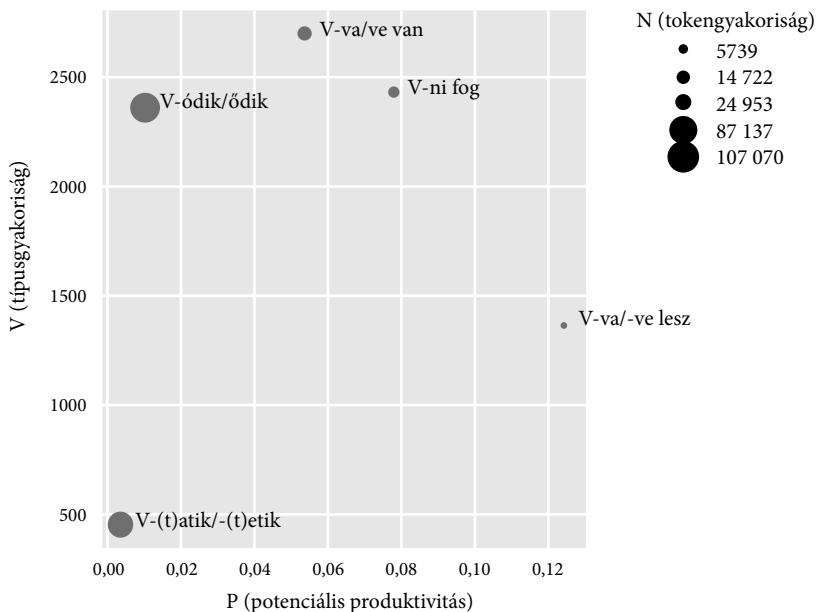


6. ábra. A vizsgált konstrukciók globális produktivitása a sajtónyelvi alkorpuszban

## V. számú melléklet: Tudományos ismeretterjesztő alkorpusz

9. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n<sub>1</sub>) és potenciális produktivitása (P) a tudományos ismeretterjesztő alkorpuszban

szerkezet	N	V	n <sub>1</sub>	P
V-ódik/-ődik	107 070	2355	1059	0,0099
V-va/-ve van	24 953	2698	1339	0,0537
V-(t)atik/-(t)etik	87 137	444	285	0,0033
V-va/-ve lesz	5739	1359	716	0,1248
V-ni fog	14 722	2427	1148	0,0780

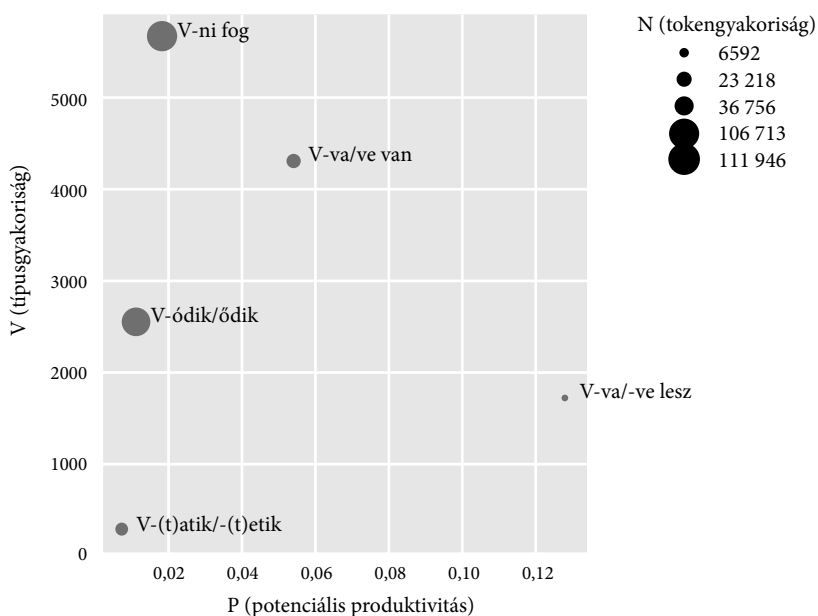


7. ábra. A vizsgált konstrukciók globális produktivitása a tudományos ismeretterjesztő alkorpuszban

## VI. számú melléklet: Beszélt nyelvi alkorpusz

10. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n1) és potenciális produktivitása (P) a beszélt nyelvi alkorpuszban

szerkezet	N	V	n <sub>1</sub>	P
V-ódik/-ődik	106 713	2537	1215	0,0114
V-va/-ve van	36 756	4321	1983	0,0540
V-(t)atik/-(t)etik	23 218	270	172	0,0074
V-va/-ve lesz	6592	1718	844	0,1280
V-ni fog	111 946	5675	2048	0,0183



8. ábra. A vizsgált konstrukciók globális produktivitása a beszélt nyelvi alkorpuszban