

# Poliszémia politópokkal

Kornai András

*SZTAKI*

andras@kornai.com

## Kivonat

A poliszémia és a homonímia elkülönítése a lexikai szematika régi megoldatlan problémája, melyet ebben a cikkben a számítógépes nyelvészetben központi szerepet játszó szövektorok politópokra való általánosításának segítségével oldunk meg.

**Kulcsszavak:** szövektor, politóp, homonímia, poliszémia

**English title:** Polysemy via polytopes

**Abstract:** Differentiating homonymy and polysemy is among the oldest and most vexing problems of lexical semantics, which we solve here by generalizing the broadly used word vector model of computational linguistics from vectors to polytopes.

**Keywords:** word vector, polytope, homonymy, polysemy

## 1. Bevezetés

A poliszémia és a homonímia elkülönítése talán a legnagyobb megoldatlan probléma, amit a korábbi strukturalista és generatív elméletek a modern lexikai szemantikára hagytak. Apresjan (1974) a szó-szintű poliszémiát úgy definiálja, mint

a szavak azon képessége, hogy számos egymástól eltérő, de egymással rokon jelentést vegyenek fel (ahol a jelentések rokonságát a definíciókban található közös komponensek szolgáltatják)<sup>1</sup>

A tankönyvek és kézikönyvek gyakran hangsúlyozzák, hogy a poliszémia gyakori jelenség: Zgusta (1971) szerint „bármely nyelv bármely szótárát kinyitva minden oldalon fogunk találni példákat a poliszémiára. A jelenség annyira gyakori, hogy általában csak nagyon specifikus szavaknál nem lép fel”.<sup>2</sup>

Ezzel szemben a homonímia csupán a hangalakok vagy írott alakok véletlen egybeesése olyan szavaknál, vagy nagyobb egységeknél, melyek jelentése nem mutat rokonságot. Ez ugyan sokkal ritkább jelenség, de pusztán a rendszerszerűség és szórványosság (kétségkívül helytálló)

---

<sup>1</sup>The usual notion of lexical polysemy [is] the capacity of a word to have several distinct but related meanings (if by relatedness we mean the presence of common components in their definitions). [p 7]

<sup>2</sup>One can open any dictionary of any language and one will find examples of polysemy on every page. Indeed, polysemy is such a frequent phenomenon that it is generally only very specific words (. . .) that are not polysemous. [p 61]

megkülönböztetése az egyedi esetekre nem ad támpontot. Mint Lyons (1977) hangsúlyozza, „a poliszémia és a homonímia elméletfüggetlen elhatárolásának bizonytalanok a kritériumai”.<sup>3</sup>

Cikkünk 2. szakaszában áttekintjük a témakör modern irodalmát, és a poliszémia/homonímia finomabb tipológiai beosztását használva megállapítjuk, hogy a Lyons által már fél évszázada is hiányolt, elméleti megalapozottságú megkülönböztetés azóta sem született. A 3. szakaszban röviden ismertetünk egy, az elméleti nyelvészek körében ma még kevésbé ismert elméletet, a *vektoros szemantikát*. Ezt a modellt a 4. szakaszban kiterjesztjük vektorokról *politópokra*, egymástól hipersíkokkal elválasztott tartományokra. Ez lehetővé teszi azt, hogy a 5. szakaszban elméleti szinten is elkülönítsük az egymással részben szomszédos (poliszém) és nem szomszédos (homonim) politóp-párokat, tehát egy olyan szomszédossági viszonyt vezessünk be, amely vektoroknál nem lenne értelmezhető. Összességében tehát a problémát újszerű elméleti alapon, a számítógépes nyelvészetben gyökerező vektoralapú szemantika egyfajta technikai módosításával oldjuk meg.

## 2. A jelenségkör finomabb felosztása

Míg a poliszémia többféleképpen is osztályozható (ha úgy tetszik, maga a fogalom is poliszém), a homonímia tekintetében az irodalom eléggé egységes: akkor beszélünk homonímiáról, „amikor egy lexikai egység véletlenül két különálló és nem rokon jelentést hordoz”.<sup>4</sup> Az egységesség valószínűleg annak is köszönhető, hogy míg a poliszémia gyakori, sőt tipikus, addig a homonímia relatíve ritka jelenség, és kevés (szórvány) adatot eleve nehéz további kategóriákra bontani.

Könnyű viszont a homonímiát és a poliszémiát egy kategória, a *többértelműség* alatt összevontan kezelni, ezt teszi már Weinreich (1964) is, aki a homonímiát *kontrasztív*, a poliszémiát pedig *komplementáris* többértelműségnek nevezi. Formailag ezt az eljárást követi Pustejovsky (1995) is, aki a homonímiával nem is nagyon foglalkozik (akcidentális jelenségről lévén szó, ez érthető is), de Weinreich eredeti példái világossá teszik, hogy ő kontrasztívnak (homonimnak) tekint olyan párokat is, amiket Pustejovsky poliszémként fog fel. Az alábbiakban gyakran fogunk egy olyan példát használni, amit Weinreich és Pustejovsky egyaránt tárgyalnak, ez a *turn* 'fordul, fordít' ige és a morfológiailag a magyarban is egyértelműen ehhez kapcsolódó *fordulat*, *fordulás*, *fordítás* főnevek, illetve *fordítva* határozó és *fordított*, *fordult* melléknevek, melyekre együtt a FORD tő származékaiként gondolhatunk.

Az első megkülönböztetés, amit a témakör elemzői gyakran megtesznek, az a szófaji kettősség ill. annak hiánya. A magyarban a morfológia ezt a FORD szóbokorban sok helyütt egyértelműsíti, de az angolban máshol is igen gyakori, pl. *divorce* 'válás/válik'; *respect* 'tisztel/tisztelet' stb. Külön érdekes ebből a szempontból az igei *turn*, melynek a magyarban intranszítív (*fordul*) és tranzitív (*fordít*) igealak egyaránt megfelelhet. Weinreich itt a tranzitív alakot tekinti primitívnek és az intranszítívát levezethetőnek, mi viszont mindkettőt ugyanabból az absztrakt FORD elemből vezetjük le úgy, hogy az *-ul* képzővel a cselekvés helyszínét (*locus*) az alanyra, az *-ít* képzővel pedig a tárgyra helyezzük. A vektoros szemantika ezt nagyban megkönnyíti azzal, hogy a függvények argumentumszámától függetlenül minden szónak ugyanaz a logikai típusa, a vektor (rendezett szám *n*-es), bár a különböző szófajokhoz tartozó vektorok a térben észrevehetően elkülönülnek (Lévai és Kornai, 2019).

Weinreich ugyan kontrasztívnek tekinti a *he turned a crank* 'megforgatta a kurbli' és a *he turned a somersault* 'bukfencet vetett' alakokat, de ezt a tárgyak különbözőségéből vezeti le: az

<sup>3</sup>The criteria for distinguishing pre-theoretically between homonymy and polysemy are uncertain. [vol 1 p 235]

<sup>4</sup>When a lexical item accidentally carries two distinct and unrelated meanings Pustejovsky, 1995, 29. old.

egyik esetben egy konkrét fizikai tárgyról van szó, míg a másikban ’egy forgással jellemzett cselekvésről’.<sup>5</sup> Ebben a felfogásban tehát a *FORD* szemantikai elemért inkább a bukfenc a felelős, a *turn* itt inkább csak egy tartalmatlan ’könnyű ige’ (*light verb*, Jespersen, 1965), ugyanúgy, mint a magyar *vet*, ami egyáltalán nem tartalmaz forgásra utaló elemet.

Az már ebből a példából is kiderül, hogy a szófaji elkülönülés nem feltétlenül tekinthető diagnosztikusnak, annál is kevésbé, mert itt abban sem egységes a szakirodalom, hogy az intranszitiv és a tranzitiv igéket ugyanabba az osztályba tartozóknak tekintsük-e vagy sem. A kérdést tovább nehezíti, hogy az egyes nyelvek egyáltalán nem ugyanott húzzák meg a szófaj-határokat (Croft, 2000, 2023). De még ha belül is maradunk egy szófajon, akkor is joggal bírálja Weinreich a Webster’s 3rd (Gove, 1961) szótárkészítési gyakorlatát, melyben a *turn* igének több mint száz aljelentését és al-aljelentését különítik el. Ezen a modern lexikográfia annyit mindenképpen változtatott, hogy külön kezeli az idiómákat (*turn a blind eye*, *turn the other cheek*, *turn the corner*, *turn up your nose*, *turn Queen’s evidence*. . .) – ezek nagy részére ugyanúgy alkalmazható Weinreich megfigyelése, hogy a *turn* bennük csak könnyű ige, pl. *turn a new page* ’új lapot nyit/kezd’; *turn up your nose* ’felhúzza az orrát, fintorog’; stb. Hasonlóan csökkenti az al(al)jelentések kijelölésének terhét az igekötős igék (*phrasal verbs*, Vincze, 2011) különvétele is: *turn sb inside out* ’kifordít’; *turn sb in* ’felad (rendőrségen)’; *turn sg in* ’bead (házi feladatot)’ stb.

Egész más irányba indult el Roman Jakobson, aki a *bachelor* egyes jelentéseit egyetlen absztrakt alapjelentés ’unfulfilled in typical male role’ segítségével ragadta meg (ld. Kornai (2019) 3. fejezet). Ha a *FORD*-hoz hasonló absztrakt jelentést rendelünk a *turn*-höz is, akkor nagyban redukálni tudjuk az (al)alesetek számát. Lehet, hogy a leginkább specifikus jelentés ’tengely körüli (ismételt) elfordulás’ lenne, és ez jól is illeszkedik az olyan példákra, mint *he turned a crank*, vagy akár *turn a figurine (on a lathe)* ’fafigurát esztergál’. De a forgó mozgásnak egyáltalán nem kell ismétlődőnek lennie, *he turned the pancake* ’átfordította a palacsintát’, sőt még a teljes 360 fokos fordulat sem szükséges. Tipikus a 180 fok, *he turned back*; és a 90 fok *he turned to the side*; de gyakran kevesebbel is megelégszünk, *he turned his head*. Az absztrakt *FORD* tehát nem tartalmaz a fordulás mértékére vonatkozó specifikációt sem az angolban, sem a magyarban.

Ezzel a Weinreich által javasolt hármas tagolás ’szó szerinti, átvitt, képletes’ (literal, shifted, and figurative) első két kategóriája automatikusan egybeesik, hiszen a konkrét jelentések értelemszerűen nemcsak a prototipikus esetben (egészen megfordul, többször) lesznek érvényesek. De mi a helyzet a képletes esettel? Ezt most a *turn* ’átváltozik’ jelentésén fogjuk illusztrálni. Ez a magyarban is megtalálható (*hidegre fordult az idő*, *fordult farkas*, *átbucskázik körtéfavá*. . .), de az angolban sokkal gyakoribb: ’to start to have a different quality’ (*turn blue*, *turn angry*, *turn traitor*); ’if milk turns, it becomes sour’. (Bár a magyarban a tej nem megfordul, hanem megszűnik, de a probléma ugyanaz, hiszen ez a jelenség szó szerinti értelemben alvásnak sem tekinthető.) Itt tehát egy másik jelentéssel van dolgunk, melyet inkább a „minőségváltozás” címszó alatt keresnénk a szótárban – ennek leírásához mindenképp szükségünk lesz egy *VÁLT* absztrakt szemémára, amibe a ruhaváltástól az őrsgé váltásig ismét sok dolognak bele kell férnie. A kérdés politópos tárgyalására majd a 4. szakaszban térünk vissza.

Összefoglalóan elmondhatjuk, hogy a Lyons által felvetett probléma még súlyosabb is annál, mint ahogy ő fogalmaz, hiszen a homonímia és a poliszémia elhatárolására nemcsak elméletfüggetlen, de még elméletfüggő kritériumok sincsenek. A baj az, hogy mind a poliszémiát, mind a homonímiát jól tudjuk kezelni elméleti módszerekkel, csak nem ugyanabban az elméleti keretben. A definíciókban található közös komponens megléte jól tárgyalható a lexikai szemantika

<sup>5</sup>an act characterized by rotary motion [p 406]

klasszikus Katz és Fodor (1963), elméletében és annak modernebb változataiban, míg a jelentések különbözősége szigorúan definiálva van a Montague nyelvtenban és annak leszármazottjaiban. Sajnos a Montague nyelvtenban elejétől fogva szereplő 'egyértelműsített nyelv' (*disambiguated language*) algoritmikus levezetését sohasem dolgozták ki, és ezen elméletnek a szójelentésre vonatkozó technikai eszköze, a jelentésosztulátumok, nem alkalmasak a finomabb megkülönböztetésekre: ha két szójelentés egyike sem következik a másikból logikailag, akkor azok különbözőek. Ennek eredményeképp pl. az  $óra_1$  'időtartam' és az  $óra_2$  'időmérő eszköz' ugyanolyan távol kerülnek egymástól, mint a *kutya* és a *macska*, hiszen sem az időmérő meglétéből nem következethetünk hatvan perces időtartamra, sem fordítva. (A helyzet még ennél is rosszabb, hiszen a *kutyából* és a *macskából* egyaránt következethetünk *állatra*, sőt *háziállatra*, míg az  $óra_2$  tárgy, az  $óra_1$  pedig fogalom, ezeknek tehát a nyilvánvaló poliszémia ellenére semmilyen közös implikatív általánosítását nem találjuk.)

### 3. Beágyazások

Az elmúlt tíz évben a számítógépes nyelvészeten általánossá vált, hogy a szavak (diszkrét szimbolikus elemek) helyett, szóvektorokat (az  $\mathbb{R}^n$   $n$ -dimenziós tér pontjait) használják. A vektorokra mindig mint az origóból induló és a pontban végződő irányított szakaszra, azaz mint *kötött vektorokra* gondolunk. Miután a kezdőpont mindig azonos, a vektorok és végpontjuk közt nem teszünk megkülönböztetést. A szavak és a (vektor vég)pontok közti leképezést szokás szó-beágyazásnak is hívni, hiszen a szavakat a térben helyezi el, oda ágyazza be. Az alkalmazásokban a tér dimenziója 200 és 800 közt szokott változni, az  $n = 300$  a leggyakoribb. A matematikai statisztika klasszikus dimenzió-csökkentő eljárásaival, elsősorban főkomponens-elemzéssel (Pearson, 1901) még a relatíve magas (pl. a GPT2 (Radford és tsai., 2019) által használt  $n = 12288$ ) dimenziószám is hatékonyan csökkenthető: a GPT4 (OpenAI, 2023) már csak 1536 dimenzióval dolgozik, és a klasszikus analógiás feladatok (öreg : fiatal = kecske : x) megoldásában 200-nál több dimenzióra nem is nagyon van szükség. A magyar HuBERT (Nemeskey, 2021) eredetileg 768 dimenziós beágyazásai is 2-300 dimenzióra való redukció után teljesítenek legjobban az analógiás feladatokon (Gedeon, 2023).

Első látásra talán meglepő lehet, hogy a dimenziószám csökkentése a teljesítmény javulásához is vezethet, de a magyarázat egyszerű: egy ilyen lépéssel gyakran épp az irreleváns adatokat, a zajt szűrjük ki. A modern beágyazások gyakran többmillió szóalakot is lefednek, de amelyek alak a korpuszban csak ritkán fordul elő, az ahhoz tartozó vektorok gyakran zajosak. A lefedettség javítását szolgálja a szónál kisebb egységek, az ún. szódarabok (*word pieces*, ld Wu és tsai. (2016)) bevonása az elemzésbe, de miután ezek ritkán esnek egybe a nyelvészeti indokolható morfémmakkal, erre itt most nem térünk ki bővebben – a morfémmák gépi tanulási kinyeréséről ld. Ács és Kornai (2020).

Természetesen a beágyazás nem véletlenszerűen történik, hanem arra törekedve, hogy 'közeli' szavaknak egymáshoz közeli pontok feleljenek meg (a szóvektorok pszichológiai realitásáról ld. Hollenstein és tsai., 2019). Hangsúlyozzuk, hogy a módszer célja a jelentések (*Sinne*), nem pedig a jelöletek (*Bedeutungen*) közti kapcsolatok szorosságának feltérképezése. Két szót akkor tekintünk egymáshoz közelinek, ha a disztribúciójuk hasonló. Az elmélet alapjait még Harris (1954) és Firth (1957) vetették meg, de e klasszikus munkák modern olvasói igencsak csalódtak lesznek, ha ezek alapján akarják a szóvektorokat megérteni. Harris műve világos, alaposan leírt elmélet, mely a későbbi (mind számítógépes, mind elméleti) nyelvészeti fejlődés alapjait vetette meg, de se

vektorokról, se együttes előfordulási statisztikáról nincs benne szó. Firth, azon túlmenően, hogy kiadta a máig népszerű jelszót ’szomszédságáról ismerszik meg a szó’ (*you shall know a word by the company it keeps*) sajnos jóval kevésbé érthető, és munkássága gyakorlatilag teljesen hatástalan maradt. Az alap gondolatot először Schütze (1993) hozta olyan algoritmikus formába, melynek alapján már effektíve ki is lehetett számolni a szóvektorokat.

Álljon egy  $w_o$  szó egy  $2k$  hosszú környezete a  $w_{-k}, w_{-k+1}, \dots, w_{-1}, \dots, w_1, w_2, \dots, w_k$  szavakból. Ha pl.  $k = 2$ , akkor a *vennének* szó egyik (korpuszban fellelhető) környezete *bizonyára, szívesen, –, némi, felvilágosítást*. Ha minden szóhoz rendelünk egy vektort, akkor ezzel egyben minden környezethez is hozzárendelhetünk egyet, pl. a környezetben előforduló szavakhoz rendelt vektorok (súlyozott) összegét. A cél úgy beállítani a vektorokat, hogy hasonló szavak egymáshoz közeli vektort kapjanak, nem-hasonló szavak pedig egymástól távolit. A feladatkitűzés csak látszólag körkörös, hiszen a két cél tulajdonképpen egyetlen mennyiség minimalizálása, ez pedig a hasonló szavak közti távolságok (négyzet)összege: ha ez sikerül, akkor egyben a környezetek hasonlósága is biztosítva van. Természetesen eközben törekedni kell arra is, hogy a nem-hasonló szavak viszont ne kerüljenek a térben közel egymáshoz: ezt *negatív mintavétel*l érjük el, tehát úgy, hogy a szavak valódi környezetei helyett random kiválasztott (más szavakhoz tartozó) környezeteket is figyelembe veszünk, a szó és az ilyen negatív környezet közti távolságot pedig maximalizáljuk – az eljárás részletes leírását ld. Goldberg és Levy (2014)-ben.

Az áttörést (Collobert és tsai., 2011) hozta meg: ebben a cikkben *ugyanazokat* a szóvektorokat használták négy jól ismert számítógépes nyelvészeti feladat, a szófajmeghatározás (*part of speech tagging*), a durva mondattani elemzés (*chunking*), a névelemfelismerés (*named entity recognition*), és a szemantikai szerepmeghatározás (*semantic role labeling*) megoldására, és mindegyik feladatban az addigi legjobb eredményeknél jobbat értek el. Miután a feladatok erősen eltérő jellegűek, nem volt megkerülhető az a következtetés, hogy a javulás annak köszönhető, hogy a vektorok a jelentésből ragadnak meg ha nem is mindent, de annyit feltétlenül, amennyi a megoldásban már segít.

Mára a szóvektorok alkalmazása általánossá vált nemcsak az információ-visszakeresés (*information retrieval*, pl. Google), a kérdés-megválaszolás (*question answering*), az attitűd-elemzés (*sentiment analysis*), az információ-kinyerés (*information extraction*), a tartalmi kivonatolás (*summarization*) és mindenfajta szöveg-elemzés területén, hanem még olyan feladatokban is, mint a beszéd- és írás-felismerés, melyek bemenete nem szöveges, hanem hangis vagy képi. Az ilyen feladatok megoldásához ugyanis nagy szükség van annak megállapítására, hogy egy adott kontextusban, pl. *bizonyára szívesen – némi felvilágosítást*, melyik szó a legvalószínűbb, a *látanak*, vagy az ahhoz nagyon hasonló *várnak*. Ez a feladat, a *statisztikai nyelvmodellezés*, a szóvektorok bevezetésével lényegesen könnyebbé vált (Jozefowicz és tsai., 2016; Nemeskey, 2020), és a modellek eredményességét (valós adatokhoz való illeszkedését) is egyre inkább ezen mérik, nem pedig a korai fejlesztésekben még nagy szerepet játszó analógiás feladatokon. A *sorozat-sorozat transzformáció* (*sequence to sequence transformation*, Sutskever és tsai., 2014) és a rögzített  $k$  kontextushossz-paramétert kikerülő *figyelemi* mechanizmus (*attention*, Vaswani és tsai., 2017) eredményezték azt, hogy a szóvektorokon alapuló modellek ma már az emberi, sőt számos tekintetben az embert meghaladó, nyelvi teljesítményre képesek (Piantadosi, 2023; Wilcox és tsai., 2022).

Fentebb már említettük, hogy a szóvektorok által hamar lehetővé vált az un. analógiás feladatok megoldása a vektorok közti paralelogramma-szabállyal:  $a : b = c : x$  feladatokat úgy oldjuk meg, hogy a  $\vec{v}(c)$  szóvektorból kivonjuk a  $\vec{v}(a)$  szóvektort, ehhez hozzáadjuk a  $\vec{v}(b)$  szóvektort, majd megkeressük az  $n$ -dimenziós térben az eredményhez legközelebbi

vektort, amely a *gida* lesz (Makrai, 2015; Mikolov és tsai., 2013). Nem lehet kérdéses, hogy ez a jelenség csakis úgy magyarázható, hogy a vektorok a jelentéshez kapcsolódnak.

A szövektoroknak köszönhetően bekövetkezett a régóta várt áttörés a gépi fordításban is. 2017-ben a Google Translate a *lepihenni melletted menedék a szívemnek* slágerszöveget még így fordította: 'shelter to lie down next to my heart', de 2020-ra már így: 'to rest beside you is a refuge for my heart'. 2023-ban a ChatGPT a következő megoldást választja: *To rest next to you is a refuge for my heart*, de ha tovább kérdezzük (This is taken from a song. Could you maintain the internal rhyme „melletted/szívemnek”?) akkor az alábbi megoldást kínálja fel: *Yes, it's possible to maintain the internal rhyme "melletted/szívemnek" in the translation while still conveying the same meaning. One possible option could be: "To rest by your side, my heart finds a hide."* (Could you find a different translation?) *Certainly! Here's another possible translation that maintains the internal rhyme: "To nestle close, by you to repose."* Mindez elképzelhetetlen lenne a jelentés megőrzése, az információ másik nyelvre való átvitele nélkül.

Ezekhez a fejleményekhez képest talán meglepő, hogy az elméleti szemantika legújabb enciklopédikus összefoglalása, az öt kötetes Gutzmann és tsai. (2021), a szövektorokról még csak említést sem tesz. A számítógépes nyelvészet tankönyvei, pl. Jurafsky és Martin (2022) újabb kiadásaikban már természetesen részletesen foglalkoznak a szövektorokkal, és kifejezetten erről szól Kornai (2023) is, ahol a politópos általánosítást először bevezettük.

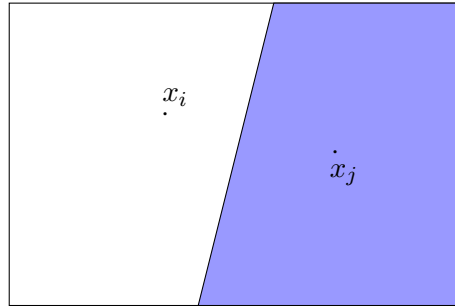
## 4. Politópok

A középiskolás geometriai anyagból az olvasónak ismerősek lesznek nemcsak a (2- és 3-dimenziós) vektorok, hanem a síkban egyenesek, térben síkok által határolt politópok is. Ezek a megszokott poligonoktól ill. poliéderektől csupán annyiban térnek el, hogy nem kell minden oldalról lehatároltnak lenniük: pl. a sík első negyede (az  $x$  és  $y$  tengelyek pozitív fele által határolt síkrész) vagy a tér első oktánsa is politópok.  $n$  dimenzióban a politópokat  $n-1$  dimenziós *hipersíkok* határolják. Egy hipersík a teret mindig két részre bontja, ezeket (affin) féltereknek hívjuk. (Az *affin* jelzőt gyakran elhagyjuk, bár szigorú értelemben csak az origón áthaladó féltereket lenne szabad féltérnek nevezni, a többi ilyenek eltolásával származtatjuk.) A politópok tehát félterek véges metszetei.

Ha adott  $k$  pont, akkor könnyen kialakíthatóak ezek körül a (Georgij Voronoj ukrán matematikusról elnevezett) *Voronoj-cellák* oly módon, hogy a tér minden pontját az eredetileg adott pontok közül ahhoz rendeljük hozzá, amelyhez a legközelebb van. Bármely két  $x_i, x_j$  ponthoz felállíthatjuk az őket összekötő szakaszra merőleges hipersíkot, mely a teret egy  $x_i$ -t tartalmazó és egy  $x_j$ -t tartalmazó féltérre bontja: előbbi  $x_j$ -t nem fogja tartalmazni, utóbbi  $x_i$ -t nem.

Mindez legjobban a kontrasztív definíciók módszerével rokonítható: a *kalap* olyan ruhadarab, amelyet a fejünkön hordunk, ellentétben a kabáttal, amit a testünkön, tehát a *ruhadarab*-ot tartalmazó féltér (vagy politóp) és a *fejlet takaró* politóp metszete. Messzire vinne, hogy a *fejletakarás*-hoz hogyan állítjuk elő a politópot (ezt Kornai (2023) 2.3 a hal-evés példáján tárgyalja). Tipikus esetben a mellékneveknek és egyváltozós predikátumoknak féltereket feleltetünk meg, ezeknél az affin eltolás mértéke a melléknév fokával hozható összefüggésbe. A főnevek megfelelői pedig inkább olyan politópok, melyeket több (a definiáló tulajdonságok által adott) féltér metszete ad ki. Ha tehát a *kutya* definíciója 'állat, ugat, hűséges', és az *állat* egy  $F_1$  féltérnek, az *ugat* egy  $F_2$  féltérnek, a *hűséges* pedig egy  $F_3$  féltérnek felel meg, akkor a *kutya* politópja  $F_1 \cap F_2 \cap F_3$  lesz.

A Voronoj-cellák fogalmát az alábbiakban némileg általánosítjuk: egy  $V = \langle \mathcal{P}, P \rangle$  **voronoid**



1. ábra. Félterek

páronként diszjunkt  $\mathcal{P} = \{Y_i\}$  politópok olyan rendszere  $\mathbb{R}^n$ -ben, melyek mindegyikéhez pontosan egy  $p_i \in P$  pont tartozik  $Y_i$  belsejében. A voronoidoknál tehát nem követeljük meg, hogy a határoló falak (hipersíkok) pontosan az egyes  $p_i$  pontok közt félúton húzódjanak. A határok pontos megválasztásának a gépi tanulásban az 1980-as évek óta nagy irodalma van, a *támvektoros gépek* (support vector machines) használata a szövegfeldolgozásban sok tekintetben megelőlegezte a mai értelemben vett szövektorokat (Joachims, 1998). Megemlítjük, hogy az elméleti alapok szorosan összefonódnak az ideghálós rendszerek tanulóképességének ún. VC-dimenziós vizsgálatával (Vapnik, 1995).

Az eredeti Voronoi-cellákat a pontokból származtatjuk, de a voronoidoknál nem követeljük meg, hogy minden cella belsejében legyenek kitüntetett pontok. A politópok határfalait más szavak, a definiáló tulajdonságok vektorai határozzák meg. A melléknevek tipikusan az 1. Ábrán láthatóhoz hasonló affin félterek: például a *kék* szóhoz tartozó  $\mathbf{k} = \vec{v}(\text{kék})$  vektor azon  $\vec{x}$ -ek halmazát jelöli ki, melyekre teljesül, hogy a  $\langle \mathbf{x}, \mathbf{k} \rangle$  skalárszorzat nagyobb egy  $b$  hasonlósági küszöbnél (*bias*), tehát a  $\mathbf{k}$  normálissal adott féltér  $\mathbf{k}$  irányba vett  $b$  hosszúsággal való eltoltjával. Ezekre a félterekre Katz és Fodor (1963) és Bolinger (1965) óta *kijelölő* (marker) néven szokás hivatkozni. A vektoros szemantika lényegéhez tartozik a típusmentesség, az hogy nem teszünk különbséget szavak és markerek, főnevek és melléknevek közt, hiszen mindkettőhöz vektorokat rendelnek a disztribúciós különbségeken alapuló algoritmusok. Amikor egy vektorhoz (affin) féltérrel rendelünk, azt használjuk ki, hogy az *állat* főnév ugyanúgy két részre osztja a teret, mint a *kék* melléknév. A legegyszerűbb (interszekatív) esetben a szó definiáló tulajdonságaihoz tartozó félterek metszete adja ki a politópot, pl.  $Y(\text{dot}) = Y(\text{mark}) \cap Y(\text{small}) \cap Y(\text{round})$  ‘a dot is a small round mark’. A nem-interszekatív melléknevekről (pl. *előző*) ld. Kornai (2023) 3.2. A politópok egymás általi meghatározottságáról itt csak annyit kell mondanunk, hogy minden egyes szótári definíció egy egyenletet ad ki, és a rendszer egészét a definíciókból adódó egyenletrendszer határozza meg.

A teljesség kedvéért érdemes azt is megemlíteni, hogy nem minden elhatárolás nyelvi jellegű: a kognitív *prototípus-elmélet* (Rosch, 1975) szerint a szavak (legalábbis a főnevek) nagy részéhez tartozik jellegzetes képi tartalom is. Egy primitív kommunikációs modellhez már ennyi is elegendő: ha a beszélő meglát (mondjuk) egy gyertyát, és ez aktiválja a fejében a  $Y(\text{gyertya})$  politópot, melyben a  $p(\text{gyertya})$  pont a kitüntetett, akkor a *gyertya* szó kiejtésével a hallgató fejében is ugyanez a politóp aktiválódik a vizuális inger nélkül is. Mindehhez még az sem kell, hogy a beszélő és a hallgató voronoidjai pontosan ugyanolyanok legyenek, elég ha ezek nagyjából ugyanazokat a térrészeket fedik le.

A jelöletek külsődleges hasonlósága (tehát a nyelven kívüli, szenzoros tényező, ld. Kornai

(2023) 1.4) áll amögött is, hogy diakrón alapon olyan sokszor találunk logikailag kontrasztív, de ennek ellenére poliszém párokat: *egér<sub>1</sub>* 'mus musculus', *egér<sub>2</sub>* 'kézi mutatóeszköz számítógéphez'; *nyereg<sub>1</sub>* 'lószerszám', *nyereg<sub>2</sub>* 'hágó'; *nap<sub>1</sub>* 'égitest', *nap<sub>2</sub>* 'időtartam' stb. Ez utóbbinál sem az absztrakt oksági viszony (az égitest megjelenése miatt van nappal) hanem az érzékszervi korreláció (világos van) az elsődleges, mint azt az egy töről származó (cognate) igék és tárgyaik kapcsolata is számos esetben mutatja Höche (2009).

Általános poliszémia-kritériumnak tekintjük azt is, ha két szójelentést a kontextus függvényében más nyelvre külön szóval fordítunk: *korsó<sub>1</sub>* 'jug', *korsó<sub>2</sub>* 'mug'. Az egynyelvű magyar beszélő esetleg észre sem veszi, hogy e két fogalom szemantikailag külön tartományt jelöl ki, pedig eltérő az anyaguk (cserép, ill. üveg), a funkciójuk (tárolás, ill. ivás), és a belekerülő folyadék is tipikusan más (víz, ill. sör). (Sokak fülében a *sír* 'cry' és a *sír* 'grave' is egy fogalom, sőt akár a *fű* és a *fa* szókezdő *f*-jében is 'növény' értelmű fonesztémát látnak.) Az viszont nem szubjektív megítélés kérdése, hanem objektív, minden magyarul és angolul egyaránt tudó beszélő számára világos nyelvi tény, hogy a *korsó* fordításánál a *jug* és a *mug* nem kicserélhetőek, még akkor sem, ha létezik egy olyan angol szó, *pitcher*, amit mindegyik fajta kontextusban jól tudunk használni.

## 5. Homonímia és poliszémia politópokkal

Ezen előkészületek után már rátérhetünk a homonímia és a poliszémia elkülönítésének mikéntjére. A tankönyvekből gyakran visszaköszönő példa az angol *bank<sub>1</sub>* 'pénzintézet' és *bank<sub>2</sub>* 'folyópart'. Hanks (2000) lexikográfiailag így különíti el a kettőt:

<i>bank<sub>1</sub></i>	<i>bank<sub>2</sub></i>
is an institution	is land
is a large building	is sloping
for storage	is long
for safekeeping	is elevated
of finance/money	situated beside water
carries out transactions	
consists of a staff of people	

Már ebből is világosan látható, hogy a két politópot számos hipersík választja el:  $Y(\textit{bank}_1)$  része az  $Y(\textit{intézmény})$  térrésznek, míg  $Y(\textit{bank}_2)$  a  $Y(\textit{föld})$  térrésznek lesz részhalmaza, márpedig e kettő diszjunkt: nincs az az intézmény, ami föld is lenne.

A generatív szintaxis kedvelt módszere, hogy olyan példákat konstruál, amik a gyakorlatban ugyan elenyésző ritkasággal fordulnak csak elő, de nyelvtani helyességükhöz nem fér kétség. Esetünkben is megkonstruálható az *I hid my money at the ...* környezet, melyről valóban ugyanúgy elképzelhető, hogy a pénzemet a folyóparton rejtettem el, mint hogy a bankban. Sőt, még olyan kontextust sem nehéz konstruálni, ahol szinte minden megjelenhet: *Csalódtam a ... -ban/ben!*. A vektoros szemantikát azonban ezek nem befolyásolják: hiába fordulnak elő a korpuszban az ilyesfajta környezetek is, bajt csak akkor okoznának, ha egyértelműsítésre alkalmas környezetek nem is lennének. A valóságban azonban a környezetek legtöbbje a kétféle jelentést világosan elkülöníti. Hanks (2000) az alábbi példákat hozza:

people without bank accounts; his bank balance; bank charges; gives written notice to the bank; in the event of a bank ceasing to conduct business; high levels of bank deposits; the bank's solvency; a bank's internal audit department; a bank loan; a bank

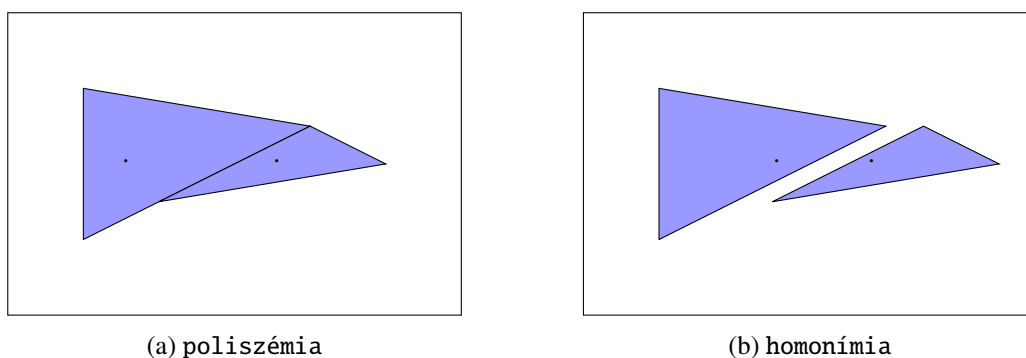


manager; commercial banks; High-Street banks; European and Japanese banks; a granny who tried to rob a bank

illetve

the grassy river bank; the northern bank of the Glen water; olive groves and sponge gardens on either bank; generations of farmers built flood banks to create arable land; many people were stranded as the river burst its banks; she slipped down the bank to the water's edge; the high banks towered on either side of us, covered in wild flowers.

Két különböző vektor távolsága mindig pozitív, de két politóp távolsága lehet nulla is, ha érintkeznek. A homonímiát és poliszémiát tehát nem az egyes jelentésekhez tartozó vektorok, hanem az e vektorokhoz tartozó politópok (voronoid-cellák) távolsága alapján különítjük el. Két jelentés **homonim**, ha a hozzájuk tartozó politópok közti távolság pozitív, és **poliszém**, ha ez a távolság nulla. A különbséget (két dimenzióban) a 2. ábrán szemléltetjük. Mint itt látható, a politópok kitüntetett pontjai a (b) panelen vannak közelebb egymáshoz, de a politópok mégis itt különülnek el (homonímia).



2. ábra. nulla, ill. pozitív távolságú politópok

Ez a definíció használatos azoknál a jelentéseknél is, ahol nem jelölhető ki egyetlen prototípus (ezeket Lakoff (1987) *sugarasnak* (*radial*) nevezi), sőt azoknál is, ahol a (lexikai) szemantika (rögzült) metaforikus kiterjesztésről beszél, mint pl. *gyertya*<sub>1</sub> 'candle' és *gyertya*<sub>2</sub> 'spark plug'. Ezekben az esetekben a két érintkező politóp uniója nem is kell, hogy konvex legyen, ilyen a helyzet a mára már funkcionálisan világosan elkülönült *bachelor*<sub>1</sub> 'bakkalaureátus' és *bachelor*<sub>2</sub> 'agglegény' esetén is.

Nem vitás, hogy az ilyen esetek nagy része valóban diakrón metafora-, ill. metonímia-rögzülési folyamat eredménye, de hangsúlyozzuk, hogy a homonímia és poliszémia elkülönítéséhez eleddig nemcsak a szinkrón, hanem a diakrón elemzés sem nyújtott elegendő alapot. Tekintsük újra az angol *bank* szót: a Concise Oxford szótár Fowler és Fowler, 1958 szerint *bank*<sub>1</sub> 'Middle English banke, from Middle French banque, from Italian banca ('counter, moneychanger's bench or table'), from Lombardic bank ('bench, counter'), from Proto-West Germanic \*banki, from Proto-Germanic \*bankiz ('bench, counter'), from Proto-Indo-European \*b<sup>h</sup>eg- ('to turn, curve, bend, bow'). Igen ám, de *bank*<sub>2</sub> 'Middle English: from Old Norse bakki, of Germanic origin; related to bench. The senses 'set of things in rows' and 'tier of oars' are from French banc, of the same ultimate origin".

A szinkrón homonim állapot tehát polisziemiából alakult ki az évszázadok során, az egyes jelentéstartományok rögzülésével, sugaras kategóriák kialakulásán, azaz új prototípusok/kitüntetett pontok megjelenésén keresztül. A teljes képhez tartozna a *bank* ige 'meghajlik, megbillen' (*the airplane banked for the turn*) és az összetételekben megjelenő *bank*-alakok, pl. az akusztikai elemzésből ismert *filter bank* elemzése is – ez utóbbinak nyilván se a pénzintézethez, se a folyóparthoz nincs köze, de annál több a 'set of things in rows' jelenéshez. Mondanunk sem kell, hogy e folyamat részleteinek megismerése csak olyan adatokon keresztül lehetséges, melyekhez a nyelvelsajátítónak nincs hozzáférése. A diakrón elemzésről tehát ugyanazt mondhatjuk, amit fentebb a szórványos/tipikus megkülönböztetés kapcsán már leírtunk: természetesen igaz, de az egyedi esetekre nézve nem ad pszichológiailag reális döntési kritériumot.

Egynyelvű beszélőnél természetesen a többnyelvű (pl. angol fordítási) adatok bevonása sem reális pszichológiailag, és kiinduló példánk, a *turn* elemzésénél sem kívánunk ezzel az eszközzel élni. A FORD és a VÁLT szemémáknak is csupán az egyszerűség kedvéért adtunk magyar nevet, a lényeg nem ez, hanem a politópok egymáshoz viszonyított helyzete. A FORD és a VÁLT metszetében van az a speciális eset, amikor a tengely körül olyan dolog forog, aminek van kitüntetett iránya (pl. az embereknél az arc, az autóknál a menetirány, a folyóknál a folyás iránya stb), hiszen ilyenkor a fordulás egyben ennek az iránynak a VÁLTOZÁSÁT is jelenti. Ez, mint logikai szükségszerűség, nyilvánvalóan az univerzális szemantika része, de az, hogy a metszetet egy adott nyelv melyik szeméma alá sorolja be, az önkényes: az angolban a *turn* jelentése számos esetben (*turn blue, turn hostile . . .*) kiterjed olyan jelenségekre, mint a bukfencet vető sáman, ahol a magyarban inkább a *válik/változik/alakul* igéket használjuk.

Összefoglalóan tehát annyit mondhatunk, hogy a homonímiát szinkrón szempontból nem tekinthetjük a polisziemia alesetének, csupán olyan véletlenszerű jelenségnek, mint az angol *ember* 'parázs' és a magyar *ember* szó hangalakjának és írásképeinek egybeesése. A polisziém (al)jelentések száma sokkal kisebb a hagyományos lexikográfia által tételezetténél (Kornai, 2023 1.4 és 2.3 ezt a *be* létige példáján illusztrálja), mert a tipikus aljelentések kontrasztjáért maga a kontrasztív kijelölő a felelős, nem pedig az (általában nagyon absztrakt) alapjelentés. Nem mondhatjuk, hogy az angol *brother* szó többértelmű csupán azért, mert a magyar vagy a kínai lexikon képesek ennek jelentéstartományát 'idősebb/fiatalabb fivér' formában felosztani, míg ugyanez a feladat az angolban csak szintaktikai parafrázis segítségével oldható meg. Annyit mondhatunk csupán, hogy az univerzális szemantikát elsajátító rendszerek (mint pl. a több mint száz nyelvet egyszerre tanuló ChatGPT) esetén érdemes külön politópokat tételezni, és az egyes nyelvekben érvényesülő polisziemiát ilyen (egymással határos) politópok nyelvfüggő rendszereivel már meg lehet ragadni. Ez a felfogás (a vektoros adaléktól eltekintve) egyáltalán nem új: már Mel'čuk, 1960 is ilyen atomokból álló szemantikát javasolt a gépi fordításhoz.

## Köszönetnyilvánítás

Köszönettel tartozunk az ÁNYT bírálóinak, akik az előző változat számos pontjára tettek nagyon hasznos megjegyzéseket. A munkát részben az EU RRF-2.3.1-21-2022-00004 támogatta a Mesterséges Intelligencia Nemzeti Laboratórium (MILAB) keretében.

## Hivatkozások

- Ács, Judit és András Kornai (2020). “The Role of Interpretable Patterns in Deep Learning for Morphology”. *XVI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2020)*. Szeged, 171–179. old.
- Apresjan, Ju. D. (1974). “Regular Polysemy”. *Linguistics* 142, 5–32. old.
- Bolinger, Dwight (1965). “The atomization of meaning”. *Language* 41.4, 555–573. old.
- Collobert, Ronan, Jason Weston, Lèon Bottou, Michael Karlen, Koray Kavukcuoglu és Pavel Kuksa (2011). “Natural Language Processing (Almost) from Scratch”. *Journal of Machine Learning Research (JMLR)*.
- Croft, William (2000). “Parts of speech as language universals and as language-particular categories”. *Approaches to the typology of word classes*. Szerk. Petra Vogel és Bernard Comrie. Mouton de Gruyter, 65–102. old.
- Croft, William (2023). “Word classes in Radical Construction Grammar”. *Oxford Handbook of Word Classes*. Szerk. Eva Van Lier. Oxford University Press. ISBN: 9780198852889.
- Firth, John R. (1957). “A synopsis of linguistic theory”. *Studies in linguistic analysis*. Blackwell, 1–32. old.
- Fowler, H.W. és F. G. Fowler (1958). *The Concise Oxford Dictionary of current English*. 4. kiad. Clarendon Press.
- Gedeon, Máté (2023). *Separation of word vectors in static and dynamic embeddings*. BSc Thesis, Budapest University of Technology és Economics.
- Goldberg, Yoav és Omer Levy (2014). “word2vec Explained: deriving Mikolov et al.’s negative-sampling word-embedding method”. *arXiv preprint arXiv:1402.3722*.
- Gove, Philip Babcock, szerk. (1961). *Webster’s Third New International Dictionary of the English Language, Unabridged*. G. & C. Merriam.
- Gutzmann, Daniel, Lisa Matthewson, Cecile Meier és Hotze Rullmann end Thomas E. Zimmerman, szerk. (2021). *The Wiley Blackwell Companion to Semantics*. Wiley-Blackwell. ISBN: 978-1-118-78831-8.
- Hanks, Patrick (2000). “Do word meanings exist”. *Computers and the Humanities*, 171–177. old.
- Harris, Zellig S. (1954). “Distributional structure”. *Word* 10.23, 146–162. old.
- Hollenstein, Nora, Antonio de la Torre, Nicolas Langer és Ce Zhang (2019. nov.). “CogniVal: A Framework for Cognitive Word Embedding Evaluation”. *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*. Hong Kong, China: Association for Computational Linguistics, 538–549. old. DOI: 10.18653/v1/K19-1050. URL: <https://aclanthology.org/K19-1050>.
- Höche, Silke (2009). *Cognate Object Constructions in English: A Cognitive-linguistic Account*. Gunter Narr Verlag. ISBN: 978 3 8233 6489 4.
- Jespersen, Otto (1965). *A Modern English grammar on historical principles*. VI. köt. London: Allen és Unwin.
- Joachims, Thorsten (1998). “Text categorization with Support Vector Machines: Learning with many relevant features”. *Proc. European Conference on Machine Learning*, 137–142. old.
- Jozefowicz, Rafal, Oriol Vinyals, Mike Schuster, Noam Shazeer és Yonghui Wu (2016). “Exploring the limits of language modeling”. *arXiv preprint arXiv:1602.02410*.
- Jurafsky, Daniel és James H. Martin (2022). *Speech and Language Processing*. 3rd edition. URL: <https://web.stanford.edu/~jurafsky/slp3/>.

- Katz, Jerrold J. és Jerry A. Fodor (1963). “The structure of a semantic theory”. *Language* 39, 170–210. old.
- Kornai, András (2019). *Semantics*. Springer Verlag. ISBN: 978-3-319-65644-1. URL: <http://kornai.com/Drafts/sem.pdf>.
- Kornai, András (2023). *Vector semantics*. Springer Verlag. URL: <http://kornai.com/Drafts/advsem.pdf>.
- Lakoff, George (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press. ISBN: 978-0-226-46803-7.
- Lévai, Dániel és András Kornai (2019. jan.). “The impact of inflection on word vectors”. *XV. Magyar Számítógépes Nyelvészeti Konferencia*.
- Lyons, John (1977). *Semantics*. London és New York: Cambridge University Press.
- Makrai, Márton (2015). “Comparison of distributed language models on medium-resourced languages”. *XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015)*. Szerk. Attila Tanács, Viktor Varga és Veronika Vincze. Szegedi Tudományegyetem Informatikai Tanszékcsoport, 22–33. old. ISBN: 978-963-306-359-0.
- Mel’čuk, Igor A. (1960). “Grammatical meanings in interlinguas for automatic translation and the concept of grammatical meaning”. *Machine Translation and Applied Linguistics* 4, 95–114. old.
- Mikolov, Tomas, Wen-tau Yih és Geoffrey Zweig (2013). “Linguistic Regularities in Continuous Space Word Representations”. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013)*. Atlanta, Georgia: Association for Computational Linguistics, 746–751. old.
- Nemeskey, Dávid Márk (2020). “Natural Language Processing Methods for Language Modeling”. Dissz. Eötvös Loránd University.
- Nemeskey, Dávid Márk (2021). “Introducing huBERT”. *XVII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2021)*. Szeged, 3–14. old.
- OpenAI (2023). *GPT-4 Technical Report*. eprint: 2303.08774. URL: <https://arxiv.org/pdf/2303.08774.pdf>.
- Pearson, Karl (1901). “LIII. On lines and planes of closest fit to systems of points in space”. *Philosophical Magazine Series* 6 2.11, 559–572. old.
- Piantadosi, Steven (2023). *Modern language models refute Chomsky’s approach to language*. URL: <https://lingbuzz.net/lingbuzz/007180>.
- Pustejovsky, James (1995). *The Generative Lexicon*. MIT Press.
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei és Ilya Sutskever (2019). “Language Models are Unsupervised Multitask Learners”. <https://github.com/openai/gpt-2>. URL: <https://d4mucfpsywv.cloudfront.net/better-language-models/language-models.pdf>.
- Rosch, Eleanor (1975). “Cognitive Representations of Semantic Categories”. *Journal of Experimental Psychology* 104.3, 192–233. old.
- Schütze, Hinrich (1993). “Word Space”. *Advances in Neural Information Processing Systems* 5. Szerk. SJ Hanson, JD Cowan és CL Giles. Morgan Kaufmann, 895–902. old.
- Sutskever, Ilya, Oriol Vinyals és Quoc V. Le (2014). “Sequence to Sequence Learning with Neural Networks”. *Proc. NIPS*. Montreal, CA, 3104–3112. old. URL: <http://arxiv.org/abs/1409.3215>.
- Vapnik, Vladimir (1995). *The nature of statistical learning theory*. Springer. ISBN: 0-387-98780-0.

- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser és Illia Polosukhin (2017). “Attention is All you Need”. *Advances in Neural Information Processing Systems 30*. Szerk. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan és R. Garnett. Curran Associates, Inc., 5998–6008. old. arXiv: 1706.03762 [cs.CL]. URL: <http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf>.
- Vincze, Veronika (2011). “Semi-Compositional Noun + Verb Constructions: Theoretical Questions and Computational Linguistic Analyses”. Dissz. University of Szeged.
- Weinreich, U (1964). “Webster’s Third: A Critique of its Semantics”. *International Journal of American Linguistics* 30, 405–409. old.
- Wilcox, Ethan, Richard Futrell és Roger Levy (2022). *Using Computational Models to Test Syntactic Learnability*. DOI: [https://doi.org/10.1162/ling\\_a\\_00491](https://doi.org/10.1162/ling_a_00491).
- Wu, Yonghui és tsai. (2016). *Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. URL: <http://arxiv.org/abs/1609.08144>.
- Zgusta, Ladislav (1971). *Manual of lexicography*. Prague: Academia.