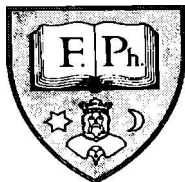


PUBLICATIONES UNIVERSITATIS MISKOLCINENSIS

SECTIO PHILOSOPHICA
TOMUS X. – FASCICULUS 3.



E TYPOGRAPHEO UNIVERSITATIS
MISKOLC 2005



TARTALOMJEGYZÉK

Csernoch Mária: Természetes nyelvi szövegek összehasonlítása elsőrendű statisztikai modellekkel	3
Dobos Csilla: A vonzat fogalmának kialakulásáról	27
Dobrovits Mihály: Étkezés a török-kori Magyarországon	49
Elterné Czöndör Klára: A zsidó-spanyol nyelv írásmódja	57
Simigné Fenyő Sarolta: A kommunikáció és a tolmácsolás modellálása	75
Szepessy Péter: A magyar holokausztot tárgyaló publikációk elemző áttekintése 1945-től a rendszerváltásig	85
Tellingner Dusán: Az etnikulturémák szerepe a műfordításban	123

TERMÉSZETES NYELVI SZÖVEGEK ÖSSZEHASONLÍTÁSA ELSŐ-RENDŰ STATISZTIKAI MODELLEKKEL

CSERNAK MÁRIA

Bevezetés

Természetes nyelvi szövegek statisztikai módszerekkel történő elemzése már a számítógépek megjelenése előtt megkezdődött, de akkor kapott igazán nagy lendületet, amikor megjelentek a nagy mennyiségű adatok viszonylag gyors feldolgozására alkalmas számítógépek. Ekkor nyílt igazán lehetőség arra, hogy a matematikai statisztika már ismert eredményeit természetes nyelvi szövegek elemzésére is használják. Napjainkra számtalan alkalmazás látott napvilágot, és önálló tudományág alakult ki a statisztika, a számítástudomány, a nyelvészet megfelelő eszközeit felhasználó részterületek összefonódásából.

Természetes nyelvi szövegek statisztikai közelítései

A természetes nyelvi szövegeket leíró statisztikai modellek számos kritérium alapján csoportosíthatók. Jelen vizsgálatok szempontjából azt a csoportosítást vettük alapul, hogy a nyelv elemeit mennyire tekintjük egymástól függőnek illetve függetlennek (Oakes 1998). Ezen szempontot figyelembe véve a következő kategorizálást lehet elvégezni:

-Nullad-rendű közelítés

Ez a legegyszerűbb modell, amelyben a szimbólumok egymástól függetlenek és egyenlő valószínűségűek (például: az egymást követő kocka dobások).

-Első-rendű közelítés

A szimbólumok függetlenek, de az előfordulási valószínűségük a szövegbeli gyakoriságokból számolható.

-Másod-rendű közelítések

Az egymást követő szimbólumok nem függetlenül kerülnek kiválasztásra, hanem a valószínűségük függ az előtte álló szimbólumtól. Ezek a közelítések adják egy nyelv bigram struktúráját, amelyben a szimbólumpárok gyakorisága annak alapján van meghatározva, hogy természetes nyelvi szövegben milyenek mérték. A Markov-modellben (Markov 1916, Arató 1970) minden egyes következő állapot csak és kizárólag a jelen állapottól függ.

-Harmad-rendű közelítések

A természetes nyelv trigram struktúrájának az előállítása lehetséges ezzel a

közelítéssel, amelyben három egymást követő elem gyakoriságát veszik figyelembe.

-n-edrendű közelítések

Az előbbieket általánosításaként beszélhetünk n -edrendű közelítésekről is, de természetes nyelvű szövegek modellezéséhez harmadrendű modelleknél magasabb számút nem szokás használni. Ekkor ugyanis már számolnunk kell az értékes kimenetek számának csökkenésével és a nagyon nagy, szerteágazó, elemző mátrixokkal, gráfokkal.

Egy összetett Markov-modellban a függőség messzebbre nyúlik vissza, egy lánc, amely megelőzi a jelen állapotot. A Markov-választás, amely a kiválasztott állapot előtti n állapotot veszi figyelembe $(n+1)$ -edrendű közelítése a vizsgált nyelvnek, amelyből az átmeneti valószínűségeket számoltuk. Ezt a modellt n -edrendű Markov-modellnak hívják.

Bármennyire is biztatónak tűnik a Markov-modellek használata, elég nagy egyetértés van abban a tekintetben, hogy a statisztikai valószínűségi modellje a természetes nyelvi kommunikációra nem alkalmazható maradéktalanul. Képtelenség megszámlálni az összes elemsorozatot egy nyelvnek. De még ha ez nem lenne is lehetetlen, a legtöbb elem előfordulása más tényezőktől függ, nem az előtte előforduló elemtől. Például a nyelvtani függőségek gyakran olyan kifejezések között állnak fenn, amelyek nem szomszédosak egymással. Nem beszélve arról, hogy a statisztikai módszer figyelmen kívül hagyja a szövegek értelmének és a diskurzusban betöltött szerepének, céljának a legtöbb tényezőjét.

(Beaugrande és Dressler 2000)

Nem szabad azonban elfelejtenünk, hogy a statisztikai modellek mellett mindig is jelen voltak és vannak az ún. determinisztikus modellek is. Mindkét megoldás mellett elhangoznak érvek és ellenérvek, hiszen egyik sem képes az emberi gondolkodás reprodukálására, így természetes nyelvi szöveg visszaadására sem. Mindennek tudatában azonban azt mondhatjuk, hogy

[...] a prabobilisztikus (valószínűségeen alapuló) modellek megfelelőbbek és valóságosabbak, mint a determinisztikus (meghatározottságon alapuló) modellek. A szerkezetépítő műveletek dinamikus jellemzése produktívabb, mint maguknak a szerkezeteknek a leírása. Arra kell törekednünk, hogy a szabályszerűségeket, stratégiákat, indítékokat, preferenciákat és alapeseteket fedezzük fel, nem pedig arra, hogy szabályokat vagy törvényeket”

(Beaugrande és Dressler 2000)

Első-rendű statikus és dinamikus statisztikai modellek

Vizsgálatainkban arra kerestük a választ, hogy egy első-rendű statisztikai modell mennyire jó közelítést adja a természetes nyelvi szövegeknek, valamint azt, hogy mivel magyarázhatóak a nyilvánvaló eltérések. Napjainkra számos olyan eredmény látott napvilágot, amelyek az eredeti szövegben előforduló szavak gyakoriságát, relatív gyakoriságát felhasználva, a szókészlet nagyságára és gazdagságára vonatkozó becsléseket képesek adni. Ezen eredmények matematikai formulák, esetleg algoritmusok formájában jelentek meg, amelyek a bemenő adatok ismeretében előállították az egyetlen kimenő értéket, tehát ún. statikus modellek voltak. A kapott eredmények feldolgozásán túl célszerű azonban ezen formulák ismeretében arra is választ keresni, hogy a modellek mennyire közelítik az eredeti szöveget, hiszen ezek az értékek megadják azt is, hogy az egyes formulák mennyire megbízhatóak. Baayen (2001) összefoglaló művéből átfogó képet kaphatunk az eddigi eredményekről és ezek esetleges pontatlanságáról is.

Korábbi kutatásaink során, szemben a statikus modellekkel, egy olyan dinamikus modellt hoztunk létre, amely a modell megépítésén túl képes modell-alapú mesterséges szövegek előállítására (Csernoch és Hunyadi 2003). A szövegek feldolgozását, a modell építését, a mesterséges szövegek előállítását, a kiértékelést, az eredeti szövegben történő visszakeresést a saját fejlesztésű, *DyMoCASAT* program végezte.

A dinamikus modell előnye a statikus modellekkel szemben, hogy a mesterséges szövegek, szemben a formulákkal, nemcsak a szókészlet nagyságára képesek megbízható eredményeket adni, hanem a szöveg a folyamatában is vizsgálható, így az is megadható, hogy a szöveg adott pontján hány új szó kerül bevezetésre. Ezek az értékek pedig összehasonlíthatók az eredeti szövegek megfelelőivel és a szövegek összevetéséből származó eredmények már felhasználhatók arra, hogy következtetéseket vonjunk le és magyarázatot adjunk az eltérésekre. A szövegek és modellek ilyen típusú elemzéséhez és az eltérések magyarázatához érdemes felhasználni azon szövegszerkezeti kutatások eredményeit, amelyek a szöveget további kisebb egységekre bontják. Ezzel a felosztással lehetőség nyílik arra, hogy a lehetséges magyarázatokat kategorizáljuk, így áttekinthetőbbé válik a probléma kezelése. Három különböző szintjét szokás a szövegnek megadni (Balázs 1985, Tolcsvai Nagy 1993, Beaugrande és Dressler 2000). Amennyiben elfogadjuk ezt a hármas felosztást, akkor az első-rendű statisztikai modellek létrehozásánál ezen három szint valamelyikén történhet meg a szavak nem-független előfordulása.

Szövegszintek

Szintaktikai vagy mondat szint (kohézió)

A mondaton belüli kohézió, amely az adott nyelv szintaktikai kötöttségeiből

adódik.

Maguk a hallott vagy látott szavak között milyen kölcsönös összefüggések vannak egy adott szósorozaton belül. A felszíni összetevők nyelvtani alakzatok és konvenciók alapján függnék egymástól, vagyis a kohézió a grammatikai függőségeken alapul.

(Beaugrande és Dressler 2000)

Szemantikai vagy bekezdés szint (koherencia)

A mondatok egymás közötti koherenciájából adódó kötöttségek: ezen a szinten a szöveget témájának azonossága teszi összefüggővé. Mindaddig, amíg a közlés elemei ugyanazon témához vagy egymáshoz szervesen illeszkedő résztémákhoz kapcsolódnak, a szöveg szemantikailag egységet alkot (Kiefer 1983, Beaugrande és Dressler 2000, Levinson 2000).

Ez a szövegvilág összetevői, vagyis a szövegfelszín alatt meghúzódó fogalmak és viszonyok kölcsönösen elérhető és releváns voltára utalnak. A fogalom esetünkben úgy határozható meg, mint valamely tudás-alakzat, amely többé-kevésbé egységesen és következetesen elérhető vagy aktivizálható a tudatban. A viszonyok azon fogalmak közötti kapcsolatot jelentik, amelyek együtt jelennek meg a szövegvilágban: minden ilyen kapcsolat magán viseli annak a fogalomnak a jelzését, amelyhez kötődik. [...] A viszonyok néha nincsenek explicitté téve a szövegben, vagyis nincsenek közvetlenül felszíni kifejezések révén. Az ember annyi viszonyt tesz hozzá az előtte álló szöveghez, amennyi csak szükséges ahhoz, hogy a szöveg értelmes legyen.

(Beaugrande és Dressler 2000)

Szöveg szint

A kohézió és a koherencia szöveg-központú fogalmak, amelyek a szövegek anyagára irányuló műveleteket jelölnék.

(Beaugrande és Dressler 2000)

Ezzel szemben szöveg szinten a szöveg, mint teljes egésznek a szerkezeti kötöttségei jelennek meg, magában foglalva a szöveg szituációs jelentését.

A szövegszerűség harmadik ismervét szándékoltságnak nevezhetjük. Ez a szöveg létrehozójának arra az igyekezetére vonatkozik, hogy a létrehozott közlés kohézióval rendelkező és koherens szöveget

alkosson, amely teljesíteni képes a létrehozójának szándékait, vagyis például ismereteket tudjon közvetíteni, vagy pedig egy tervben meghatározott célt tudjon elérni.

(Beaugrande és Dressler 2000)

Ha azonban figyelembe vesszük, hogy ezek a szintek nem mindig határolhatók el egyértelműen, akkor az is megengedett, hogy elfogadjuk a következő véleményeket: „a mondat és a bekezdés szint egymástól nehezen szétválasztható fogalom. Mindkettőnek megvan a maga határoló jele, ami megkülönböztetné őket, de ettől többről van szó, hiszen a szintaktikai formák szerveződésének elsősorban szemantikai okai vannak” (Dobi 2002). A másik oldalról viszont az is igaz, hogy „a szintaktikai kötöttségek túl nyúlnak a mondat határain, de viszonylag ritka azoknak az eseteknek a száma, amikor a bekezdés határt is átlépik előre és visszamutató hivatkozások” (Kiefer 1983).

Első-rendű statisztikai modellek lehetőségei

Az első-rendű statisztikai modelleknek nem lehet célja az eredeti mű visszaállítása, hiszen a szóalakok gyakoriságának ismerete ehhez nem nyújt elegendő információt. Ezzel szemben ezek a modellek alkalmasak lehetnek arra, hogy segítségükkel a szöveg egy-egy speciális tulajdonságát le tudjuk írni, valamint, hogy több ilyen paraméter összevetésével a szöveg egészére vonatkozó állításokat tudjunk megfogalmazni.

Első pillanatra azt lehetne gondolni, hogy egy a szavak függetlenségét feltételező (*randomness assumption*) modell azért nem képes visszaadni a szóalakok eredeti eloszlását, mert az író követte a szintaktikai és szemantikai szabályokat, míg a véletlenszerű válogatás ezt nem képes megtenni. Ha azonban nemcsak felületes szemlélőként nézzük a problémát, akkor kiderül, hogy az eredeti szöveg és a modell közötti eltérést nem ezek a megkötések okozzák, hanem a szöveg szinten bekövetkező változások. Baayen megmutatta (1996, 2001), hogy valóban nem a mondat szinten jelenlévő megkötések a felelősek az eltérésért, ugyanakkor nem vette figyelembe a szintaktikai és szemantikai eszközök kölcsönös egymásra hatását, amely már eleve kizárja, hogy a bekezdés szinten zajló események meghatározzák az eredeti szöveg és a modell közötti eltérést, így a bekezdés és a szöveg szinten történő változásokra egy időben kereste a megoldást.

A dinamikus modell alapján előállított mesterséges szövegek és az eredeti, természetes nyelven írt szöveg összehasonlítása (Csernoch 2004) már azt mutatta, hogy a szöveg szinten bekövetkező változásokkal magyarázható az eltérés. Ennek további bizonyítására az eredeti szöveg és annak fordításainak összehasonlítását végeztük el.

Módszerek

Dinamikus modell alkalmazása

A szöveget továbbra is folyamatában próbáltuk vizsgálni, amelyhez a korábban megépített dinamikus modellt (valamint a modell építését végző *DyMoCASAT* programot) használtuk. Ezt úgy végeztük el, hogy felosztottuk az éppen aktuális szöveget adott hosszúságú intervallumokra, blokkokra, és megszámloltuk, hogy hány új szóalak jelent meg az egyes blokkokban. Új szó alatt olyan szóalakokat értünk, amely a szövegben korábban még nem szerepelt, ez az első megjelenése. Az így kapott értékeket (f) grafikonon ábrázoltuk, majd elvégeztük a függvény szükséges simítását (fp). A következő lépésben, hasonlóan az eddigiekhez, vettük száz mesterséges szöveg újonnan bevezetett szóalakjainak az átlagát és az így kapott függvényt (F) kivontuk a simított függvényből ($fp-F$). A függvények különbsége egyértelműen megadta az eredeti szövegnek azokat a szeleteit, ahol az eltért a mesterséges szövegtől. Azokat az eltéréseket tekintettük szignifikánsnak, amelyek a különbségek átlagától (M) legalább a szórás (σ) kétszeresével tértek el (1-9. ábra).

Az általunk épített dinamikus modell nem korlátozódott angol nyelvű szövegek feldolgozására, hanem alkalmas különböző nyelveken írt szövegek elemzésére is. Ennek köszönhetően a fentebb ismertetett módszerrel egy szövegnek különböző nyelvű fordításait is össze tudtuk vetni az eredetivel. A fordításokkal elvégeztük ugyanazt a modellépítést, amit az eredeti szövegek esetén, majd ugyanúgy meghatároztuk a szöveg és a modellek átlagából az újonnan bevezetésre kerülő szóalakok számának a különbségét blokkonként. Az így kapott eltéréseket kell ezután az eredeti nyelven írt szöveg eltéréseivel összevetni. Ehhez ismételt a modellt építő programot használtuk. A program visszakereste azokat a szövegrészeket, ahol szignifikáns eltérést tapasztaltunk. Ezeket a szövegrészeket kellett összehasonlítani, megnézni, hogy a különböző nyelveken írt szövegekből ugyanazokat a szeleteket kaptuk-e vissza.

Felhasznált szövegek

A művek kiválasztását nagyban befolyásolta, hogy melyek azok, amelyeknek létezik és elérhető a fordítása, legalább nyomtatott formában, valamint az, hogy a kiválasztott nyelvek szerkezetükben eltérőek legyenek. Ennek megfelelően angol, magyar és német nyelvű szövegek feldolgozását végeztük el. Aszerint csoportosítva ugyanis, hogy a morfémákból hogyan képzik a nyelv a szóalakokat a három nyelv három különböző kategóriába sorolható. A német a flektáló, a magyar az agglutináló nyelvek csoportjába, míg az angol leginkább az izoláló nyelvek kategóriájába tartozik, de amiatt, hogy több különböző kategória eszközeit is felhasználja, így igazán egyikbe sem illik bele (Prószéky 1989, O'Grady 1993, Quirk et al. 1995, Uzonyi 1996, É. Kiss 1998, Kiefer 1998, Kugler 2000, Laczkó

2000).

A választás egy

- magyar regényre (Kertész Imre: *Sorstalanság*) és annak
- o német (*Roman eines Schicksallosen*¹) és
 - o angol nyelvű fordításaira (*Fateless*²),
- egy angol novellás-kötetre (Rudyard Kipling: *The Jungle Books*) és annak
- o magyar fordítására (*A dzsungel könyve*³), valamint
- két angol meseregényre (Lewis Caroll: *Alice's Adventures in Wonderland és Through the Looking Glass*, mostantól *Alice*)
- o és ezek magyar fordításaira (*Alice Csodaországban*⁴ és *Alice Tükörországban*⁵) esett.

Az 1. táblázat értékei mutatják, hogy az egyes nyelvek sajátosságaiból, valamint a fordításból adódóan a szövegszók (N), a különböző szóalakok ($V(N)$) és az egyszer előforduló szavak száma ($V(1,N)$) között lényeges eltérések mutatkoznak az egymásnak megfelelő szövegek esetén. A szintaktikai és szemantikai szinten bekövetkező változások elemzésére ez a módszer tehát nem lehet alkalmas, abban az esetben viszont, ha a szignifikáns eltérések a szöveg szinten következnek be, akkor állításunk bizonyítást nyer.

Eredmények

Sorstalanság

Kertész Imre *Sorstalanság* című művének elemzésekor hat olyan szakaszt találtunk, amelyekben hirtelen megemelkedik az újonnan bevezetésre kerülő szóalakok száma (1. ábra, 2. táblázat). Elsőként ezt a hat szignifikánsnak tekinthető eltérést vizsgáltuk meg részletesen. Azt tapasztaltuk, hogy a kiugrások valamennyien olyan helyen fordultak elő, ahol a szöveghez szervesen nem kapcsolódó, a korábbi eseményektől függetleníthető hosszabb lélegzetvételű leírás jelent meg a szövegben. A hat szövegrész, a megjelenés sorrendjében, a következő volt: megérkezés a koncentrációs táborba, megérkezés a második táborba, reggeli

¹ Aus dem Ungarischen von Christina Viragh, Rowohlt, Hamburg 1996.

² Translated by Christopher C. Wilson and Katharina M. Wilson, Hydra Books, Northwestern University Press, Evanston/Ill. 1992.

³ A fordítás a Macmillan Kiadó 1930. évi kiadásából készült. Benedek Marcell fordítása, a verseket Weöres Sándor fordította. Móra Könyvkiadó, Budapest 1976.

⁴ Fordította Kosztolányi Dezső, a fordítást az eredetivel egybevetve átdolgozta Szobotka Tibor. Móra Könyvkiadó, Budapest 1974.

⁵ Fordította Révbíró Tamás, a versbetéteket Tótfalusi István fordította, Móra Könyvkiadó, Budapest 1980.

események és az üzem leírása, kórház leírása, Pjetyka főz a kórházban, hazaindulás énekekkel.

A német nyelvű szöveg újonnan bevezetett szóalakjainak vizsgálatakor hét hirtelen emelkedést találtunk (2. ábra, 2. táblázat). Ezek közül öt teljesen egybeesett a magyar szöveg megfelelő szövegrészeivel, míg kettő a magyar szövegben nem jelent meg szignifikáns eltérésként a modellhez képest, és egyetlen olyan hely van, amely, szemben a magyar szöveggel, a németben nem jelent meg. A hét kiugrás közül az első, amelyik nem jelent meg a magyar szövegben, a vonatraszállást írja le, mely olyan esemény, ami időben megelőzi az első magyar kiugrást, a táborba érkezést.

A német szöveg második és a harmadik kiugrása ugyanannál a szövegrésznél következett be, mint a magyar szövegben, tehát megérkezés a második koncentrációs táborba, valamint a reggeli készülődés és az üzem leírása. A német szövegben akkor jelenik meg a negyedik kiugrás, amikor a főszereplő pillanatnyi lelkiállapotáról következik egy leírás.

Végül az utolsó három kiugrás újra teljes egészében megegyezik a magyar szöveg kiugrásaival. A német szöveg utolsó kiugrása még éppen az elfogadhatósági intervallumon belül esik, de ezt szignifikánsnak tekintettük. Azért tettük ezt, mert visszakeresve a szövegrészeket egy olyan éneket találtunk, amely a magyar szövegben szótagolva szerepel, míg a német szövegben a szavak egybe vannak írva.

A mindkét szövegben szignifikáns eltéréseket okozó szövegrészek összehasonlítása után megvizsgáltuk azokat a pontokat, amelyek az egyik nyelven okoztak kiugrásokat a görbén, míg a másikon nem. Először az egyetlen olyan helyet vizsgáltuk, amely a magyarban megjelent, de a németben nem. Azt találtuk, hogy a magyar szövegnek ez az első kiugrása szignifikánsan ugyan nem, de egy a szignifikancia-szint alatti csúccsal megjelenik a német szövegben is. A német szövegben azért nem kaptunk újabb kiugrást a táborba érkezéskor, mert a vonatraszállás, a vonat leírására használt szavak nagyban fedik a tábor jellemzésére használt szavakat.

A másik típusú két kiugrás, amelyik a német szövegben igen, de a magyarban nem volt jelen, eltérően viselkedett. A német szöveg 209. blokkjánál megfigyelhető kiugrás (n209, 2. táblázat) egyáltalán nem jelent meg a magyar szövegben. A másik ilyen kiugrás, n518, a főszereplő lelkiállapotának leírását tartalmazza. Az n518-as csúcshoz tartozó német szövegrészt visszakeresve a magyar grafikonon és szövegben megtaláltuk a megfelelő szignifikancia-szint alatti csúcst (m402).

Így, a magyar és a német szöveget összehasonlítva, összességében nyolc helyen teljes egybeesést tapasztaltunk, ezzel szemben kizárólag egyetlen olyan szövegrészt találtunk, amely csak a német szövegben okozott emelkedést az újonnan bevezetett szóalakok számában.

Ez az összehasonlítás azt mutatja, hogy két, szerkezetében eltérő nyelvű szövegben ugyanazokon a helyeken emelkedett meg az újonnan bevezetett szavak száma. Olyan szövegrészeket találtunk ezek mögött, amelyek hosszabb leírásokat tartalmaznak, a szöveg teljes egészéhez marginálisan kapcsolódva. A természetes nyelvi szövegeket összehasonlítva a modell által előállított mesterséges szövegekkel arra a következtetésre jutottunk, hogy az általunk használt első-rendű statisztikai modell valóban csak szöveg szinten bekövetkező változásokban tér el az eredeti szövegtől, azokat nem képes visszaadni. Ha ez nem így lenne, akkor a magyar és a német szöveg összehasonlítása nem eredményezhette volna az összeeséseket a mondat és bekezdés szinten meglévő megkötések miatt.

Az angol szöveg elemzésekor is hasonló eredményeket kaptunk (3. ábra, 2. táblázat). Olyan helyeken jelentkeztek a görbén kiugrások, ahol a műbe egy hosszabb lélegzetű leírás került. Ezek nagy része most is megegyezett a magyar és a német szöveg kiugrásaival. Annyiban történt változás, hogy az angol szövegben jelent meg a legtöbb, összesen nyolc, csúcs, amely lényeges eltérésnek tekinthető a szöveg megszokott menetéhez, valamint a modellhez képest.

Az angol szöveg elején megjelent három kiugrás (a43, a71, a156, 2. táblázat), amely sem a magyar, sem a német szövegben nem szerepelt, de mind egy-egy részletes leírást ad. A középső három kiugrás megegyezik a másik két szöveg kiugrásaival, míg a két utolsó olyan leírás, amely csak az angol szövegben okozott szignifikáns eltérést, de jellegét tekintve ezek is hasonlóak az előzőekhez: valamiféle, a szöveg egészét tekintve váratlan leírás jelent meg a műben.

Az angol nyelvű szöveget összehasonlítva a magyar és a német szöveggel az a447-es kiugrás, tehát az utolsó előtti az angolban, mind a magyar, mind a német szövegben jelen van, de értékük a szignifikancia küszöb alatt marad, míg az a618-as, az utolsó, megfelelője a magyar szövegben ott van, de a németben nem okozott kiugrást. Az a165-ös csúcs a vonat indulása körüli eseményeket írja le, míg ennek német megfelelője a vonat egy későbbi indulásánál jelent meg. Fordítva, az angol szövegből hiányzó négy kiugrás közül kettő (a511, és az a650) beazonosítható volt, míg a másik kettő nem (m429 = n552 és m459 = n587).

A három szöveget összehasonlítva azt tapasztaljuk, hogy a magyar szövegben megjelenő első kiugrást, a német szövegben egy, az angolban pedig három is megelőzte. Érdeemes elgondolkozni azon (és ez további kutatásokat, feldolgozásokat tesz szükségessé), hogy mi okozhatja ezt a jelenséget.

Az angol nyelvű szöveg elemzésénél választ kaptunk arra is, hogy a fejezethatároknak mennyire van meghatározó szerepe az újonnan bevezetett szóalakok számának az emelkedésében. Számos olyan vélemény fogalmazódott meg korábban, hogy egy új fejezet indítása a szókészlet megváltozásával jár együtt. Korábbi vizsgálataink során kapott eredmények cáfolták ezeket a feltételezéseket, mivel sem a fejezethatároknak, sem novellák összefüzésekor az új novellák kezdetén nem kaptunk szóalakszám emelkedést, még akkor sem, ha azok

különböző szerzőtől származtak. Csak és kizárólag abban az esetben, ha egy hosszabb lélegzetű leírás, esetleg stílusváltás történt meg a fejezet kezdetén (Csernoch 2004).

Összehasonlítva a *Sorstalanságot* annak német és angol fordításával, ugyanezt tapasztaltuk. Nem a fejezethatárokon jelent meg nagy mennyiségű új szóalak. Az angol szövegben ugyanis nem ott vannak a fejezethatárok, mint a magyarban és a németben. Ha a fejezethatároknak az is a funkciója lenne, hogy megemeljék a szóalakok számát, akkor az angol nyelvű szövegben egészen más helyeken kellett volna megjelenjenek a kiugrások, mint a másik két nyelven. Ez pedig nem így történt.

A „The Jungle Books” és „A dzsungel könyve” összehasonlítása

Vizsgálatainknak ebben a szakaszában Kipling: *The Jungle Books* című művének teljes, első és második kötetét is tartalmazó szövegét dolgoztuk fel. Már az első kötet elemzésénél is láttuk (Csernoch 2004), hogy a modell és az eredeti szöveg közötti szignifikáns eltéréseket azok a leírások eredményezték, amelyek a történethez csak marginálisan kapcsolódnak, és inkább formai, hangulati, mintsem tartalmi, a szöveg megértéséhez nélkülözhetetlen szerepet töltenek be.

Az eredeti novelláskötetet összehasonlítva a magyar fordítással, az első figyelemre méltó eltérés, hogy nem egyezik az elbeszélések sorrendje. A magyar fordításban nem tartották meg az eredeti sorrendet, hanem előre kerültek a dzsungelben játszódó, és a könyv végére a más helyszínű történetek. A történetek relatív sorrendje egy-egy eltéréstől eltekintve megegyezik az eredetivel (3. táblázat).

A 4. és a 7. ábrák az eredeti angol, míg az 5. és a 6. ábrák a magyar sorrendű szövegek alapján készültek. Az angol sorrendű ábrákon a négyzetek (□) jelzik azokat a helyeket, ahol vagy az angol vagy a magyar nyelvű szövegben szignifikáns eltérést találunk az újonnan bevezetett szóalakok számában. A magyar sorrendű ábrákon a körök (○) jelzik ugyanezeket a pontokat. Az ábrákon megjelölt helyek sorszámát a megfelelő táblázatban az első oszlop számai adják. Összevetve a 4. és a 7. ábra, valamint az ábrákhoz tartozó 4. és 7. táblázat értékeit, összesen tíz olyan szövegrészt találtunk, amelyik vagy az angol vagy a magyar nyelvű szövegben szignifikáns eltérést okozott. A tíz pont közül kilenc olyan, hogy annak megfelelője a másik nyelven is elérte vagy megközelítette a szignifikancia szintet. Egyetlen olyan pontot találtunk, ez az angol szövegben a legelső, amely nem jelent meg a magyarban. Hasonló pontokat találtunk a *Sorstalanság* és fordításainak összehasonlításánál is. A magyar szövegben később jelenik meg az első szignifikáns eltérést jelentő pont, mint az angolban.

A magyar sorrend szerint rendezve az angol nyelvű meséket is, ismételten tíz olyan helyet találtunk, amely legalább az egyik nyelven elérte a szignifikancia-szintet. Hasonlóan az eddig tapasztaltakhoz, a magyar szövegben később jelent

meg az első eltérés, mint az angolban. Mivel a magyar fordítás nem a megjelenés sorrendjét, hanem valamiféle logikai sorrendet követ, így nem meglepő, hogy a *King's Ankus* sokkal nagyobb kiugrást eredményezett, mint eredetileg. Az eredeti sorrendet tartva ugyanis a *King's Ankus* történetét már megelőzte egy emberekkel kapcsolatos történet, *The Miracle of Purun Bhagat*. A magyar sorrendnél a szórás ($\sigma = 3,1866$) is nagyobb, mint az angolnál ($\sigma = 3,0275$). A magyar sorrendnél sokkal kiegyensúlyozottabb, logikusabb a novellák elrendeződése, így egy a sorba nem illő novella sokkal nagyobb kiugrást eredményezhet.

A mesék sorrendjétől függetlenül is azt tapasztaltuk, hogy a kiugrások azokon a helyeken jelentek meg, ahol a szöveg stílusától eltérő hosszabb leírás, felsorolás jelent meg a műben. Ezek többsége, különösen a jelentősebb kiugrások mind a négy szövegben megjelentek mint szignifikáns eltérések, függetlenül nyelvtől, sorrendtől.

Az „Alice's Adventures in Wonderland” és a „Through the Looking Glass” elemzése

Azért esett a választás az *Alice*-történetekre, mert Petőfi S. János (1990) ezek egy részletes elemzését adja, összevetve az eredeti angol szöveget és annak fordítását. Arra voltunk kíváncsiak, hogy a Petőfi által megfogalmazott, a szóhasználatra vonatkozó szubjektív vélemények mennyiben támaszthatók alá a számítógépes feldolgozás eredményeivel. (Itt csak azokat az észrevételeket említem, amelyek vizsgálataink szempontjából érdekesekek.) Petőfi úgy ítéli meg, hogy a magyar fordítások nem követik hűen az eredeti szöveget sem szöveg-, sem képanyagban. Ezen túl azt találta, hogy vannak a műben olyan fejezetek, amelyek között az átmenetek sokkal gördülékenyebbek, mint ahogy az egy fejezethatáron várható volna. Ezen, a fejezetek közötti szokatlan átmeneteket rendhagyó formai elemekkel jelzi a szerző.

Hasonlóan más művekhez, azt találtuk, hogy akkor emelkedett meg az újonnan bevezetett szóalakok száma, amikor a szövegben olyan szövegrész jelenik meg, ami eltér a mű stílusától. Ezeknél a műveknél nem mondhatjuk, hogy a versek, amelyeknél megnövekedett a szóalakok száma csak kiegészítő szerepet töltenek be műben, itt a stílusváltásnak van meghatározó szerepe. A fejezethatárok viszont abban az esetben sem hoztak látványos szóalakszám emelkedést, amikor a hagyományos módon kapcsolódtak egymáshoz. Így tehát a formabontó fejezet összecsatolások nem eredményezték a szóalakok számának változását, azok rendhagyó viselkedését. Továbbra is igaz, hogy a fejezethatárok, önmagukban a fejezethatár megjelölése, nem jár együtt az újonnan bevezetett szóalakok számának emelkedésével.

A nagyobb kiugrások mind a két nyelven megjelentek ugyanazokon a helyeken. A magyar szövegben találtunk a szöveg végén egy kiugrást, ami hiányzott az angoltól, ez várhatóan a fordító szóhasználatának következménye. Az

angol szövegben találtunk négy olyan kiugrást (a93, a120, a171, a203, 8. és 9. ábra, 8. táblázat), amely a magyar szövegben nem jelent meg. Ezeket a kiugrásokat elemezve azt tapasztaltuk, hogy nagyon rövid intervallumon jelentek meg, maximum három blokkot érintve, és a méretük is jelentéktelen, alig érik el az $M + 2\sigma$ értéket. Úgy gondolom, hogy kétféle magyarázat is adható az angol és magyar szöveg közötti az eltérésre. Az egyik, hogy szövegnek a fordítása ezen a darabon tényleg nem adja vissza az eredeti művet, míg a másik lehetséges magyarázat, hogy a magyar szöveg zajosabb, mint az angol, a magyar agglutináló tulajdonsága miatt. Ennek következtében a rövid intervallumra kiterjedő, alig jelentős változások a magyar szövegben eltűnnek.

Az *Alice*-történetek angol nyelvű szövegének első kiugrása, az eddigiektől eltérően, megjelent a magyar szövegben is. Ez azzal magyarázható, hogy „Hódító Vilmos száraz története” annyira eltér stílusában a mű egészétől, hogy még a zajosabb magyar szövegben is okozott egy csúcsot, még ha az az elfogadhatósági tartományon belül is esik.

Összegzés

Kutatásaink során irodalmi művekben és azok fordításaiban az újonnan bevezetett szóalakok megjelenését vizsgáltuk. Egyrészt arra kerestük a választ, hogy mi indokolja a szóalakok számának hirtelen emelkedését, mikor használ az író nagy mennyiségű olyan szóalakot, amely korábban nem szerepelt a műben. Másfelől viszont azt vizsgáltuk, hogy a fordításokat összehasonlítva milyen szabályszerűségeket fedezhetünk fel a szóalakok számának változásában.

A vizsgálatok elvégzéséhez megépítettünk egy olyan dinamikus első-rendű statisztikai modellt, amely több különböző nyelven írt szöveg elemzésére alkalmasnak bizonyult, így a feldolgozásban résztvevő angol, magyar és német nyelvű szövegeket egységesen tudtuk kezelni, függetlenül a nyelvek sajátosságaitól. Az eredeti, természetes nyelvi szövegeket alapul véve a modell alapján generált mesterséges szövegeket hoztunk létre és ezeket a szövegeket, pontosabban ezen szövegek újonnan bevezetett szóalakjait, hasonlítottuk össze az eredeti szöveg megfelelő paramétereivel.

Korábbi vizsgálataink eredményei is azt mutatták, hogy az eredeti szövegben akkor emelkedett meg az újonnan bevezetett szóalakok száma, amikor az író egy hosszabb lélegzetű, a mű stílusától eltérő szövegrészt illesztett be. Ezek a helyek egybeestek az eredeti és a mesterséges szövegek összehasonlítása során kapott eltérésekkel, jelezve, hogy az eredeti szöveg és a modell közötti eltérések a szöveg szinten bekövetkező változásokkal magyarázhatóak. Ennek az állításnak a további bizonyítására az eredeti szöveg és annak fordításainak az összehasonlítását végeztük el. Azt vártuk, hogy ha a modell és az eredeti szöveg közötti eltérések valóban szöveg szinten végbemenő változások eredményei, akkor lényegtelen, hogy milyen nyelven íródott a szöveg, milyen szintaktikai és szemantikai

megkötések kényszerítették az írókat a szövegvezetésre.

Az összehasonlításokat elvégezve, a várakozásoknak megfelelően, azt tapasztaltuk, hogy függetlenül a nyelv sajátosságaitól és a fordítói szabadságtól, az újonnan bevezetett szóalakok száma közel ugyanazokon a helyeken emelkedett meg. Ezek a helyek szinte kivétel nélkül szignifikáns vagy közel szignifikáns eltérést okoztak mind az eredeti műben, mind pedig a fordításokban.

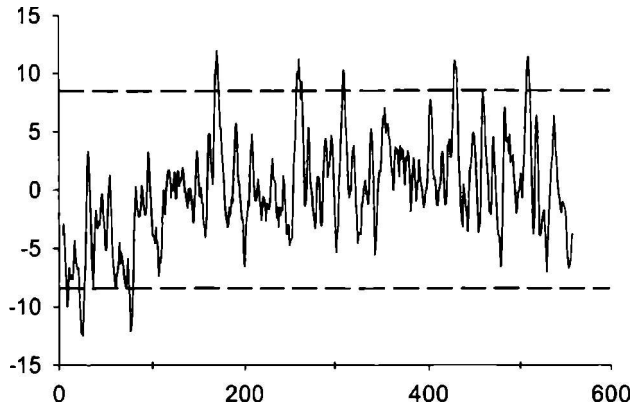
Azt a néhány helyet, ahol nem találtunk egyezést, további elemzésnek, feldolgozásnak érdemes alávétetni, hiszen a fenti eredmények azt mutatják, hogy a változások szöveg szinten következnek be. Ha az eredeti műben vagy a fordításban nem jelenik meg ez a változás, akkor egyrészt érdemes az egyes nyelvek sajátosságait figyelembe véve a modell további finomításait elvégezni. Ezzel a korrekcióval várhatóan kiküszöbölhető, hogy a magyar szövegekben később jelenik meg az első kiugrás, mint a többi nyelven. Másrészt viszont az itt ismertetett eljárás felhasználható a fordítások értékelésére, annak elemzésére tehát, hogy egy fordítás mennyire hűen adja vissza az eredeti művet.

Irodalom

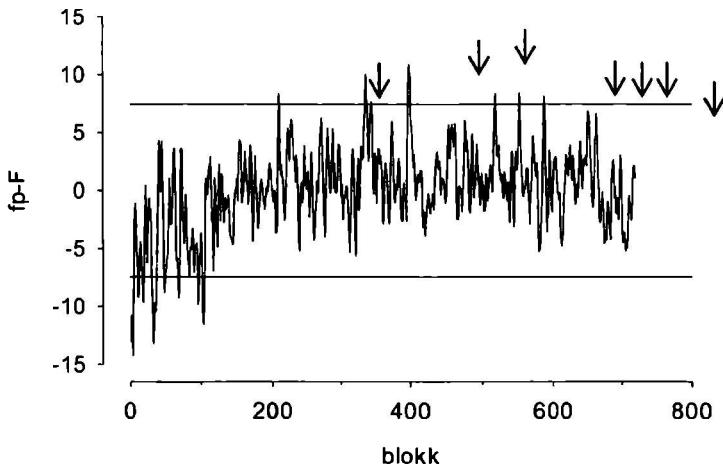
- Arató M. Knuth E. 1970: *Sztocasztikus folyamatok elemei*, Tankönyvkiadó, Budapest
- Baayen, R. H. 1996: „The effect of lexical specialization on the growth curve of the vocabulary”, *Computational Linguistics* 22: 455-480.
- Baayen, R. H. 2001: *Word Frequency Distributions*, Kluwer, Dordrecht
- Balázs J. (1985): *A szöveg*, Gondolat, Budapest
- Beaugrande, R. Dressler, W. 2000 (1981): *Bevezetés a szövegnyelvészetbe*, Corvina, Budapest
- Csernoch M. Hunyadi L. 2003: „Szótípusok bevezetésének szabályszerűsége magyar és angol nyelvű nyomtatott szövegekben”, *Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, 24-30.
- Csernoch M. 2004: „Another method to analyze the introduction of word-types in literary works and textbooks”, *The 16th Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities*, Göteborg University, 44-45.
- Dobi E. 2002: „A pragmatika szerepe a nyelvi rendszer egységeinek leírásában”, in Andor J. – Benkes Zs. – Bókay A. (szerk.): *Szöveg az egész világ. Petőfi S. János 70. születésnapjára*, Tinta, Budapest
- É. Kiss K. 1998: „Mondattan”, in É. Kiss K. – Kiefer F. – Siptár P. (szerk.): *Új magyar nyelvtan*, Osiris, Budapest
- Kiefer F. 1983: *Az előfeltevések elmélete*, Akadémiai, Budapest
- Kiefer F. 1998: „Alaktan”, in É. Kiss K. – Kiefer F. – Siptár P. (szerk.): *Új magyar*

nyelvtan, Osiris, Budapest

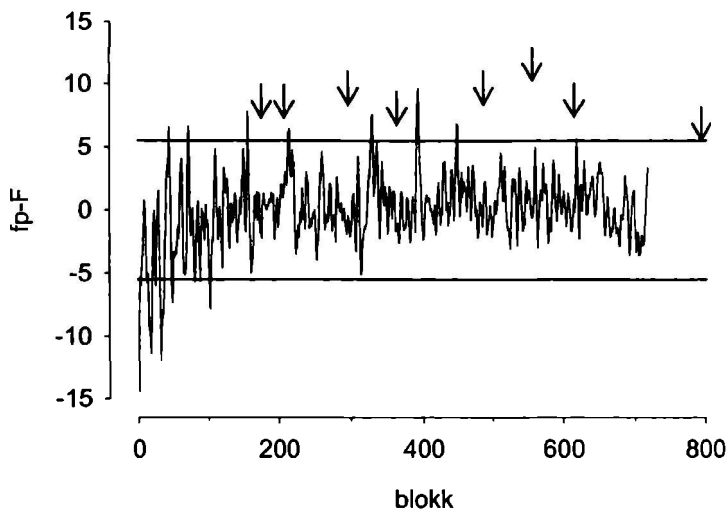
- Kugler N. 2000: „Alaktan”, in Balogh J. – Haader L. – Keszler B. – Kugler N. Laczkó K. Lengyel K. (szerk.): *Magyar grammatika*, Nemzeti Tankönyvkiadó, Budapest
- Levinson, S. C. 2000: *Presumptive Meanings. The Theory of Generalized Conversational Implicature*, A Bradford Book, The MIT Press, Cambridge/Mass. – London
- Markov, A. A. 1916: „An application of statistical method”, *Izvestiya Imperialisticheskoy Akademii Nauk*, 6 (4): 281-297.
- Oakes, M. P. 1998: *Statistics for Corpus Linguistics*, Edinburgh University Press
- O’Grady, W. – Dobrovolsky, M. – Aronoff, M. 1993: *Contemporary Linguistics. An Introduction*, New York: St. Martin’s Press
- Petőfi S. J. 1990: *Szöveg, szövegtan, műelemzés*, Országos Pedagógiai Intézet, Budapest
- Prószéky G. 1989: *Számítógépes nyelvészet*, Számítástechnika-Alkalmazási Vállalat, Budapest
- Quirk, R. – Greenbaum, S. Leech, G. – Svartvik, J. 1995: *A Comprehensive Grammar of the English Language*, Longman Group UK Limited, London – New York
- Tolcsvai Nagy G. 1993: *A szövegek világa*, Nemzeti Tankönyvkiadó, Budapest
- Uzonyi P. 1996: *Rendszeres német nyelvtan*, Aula, Budapest



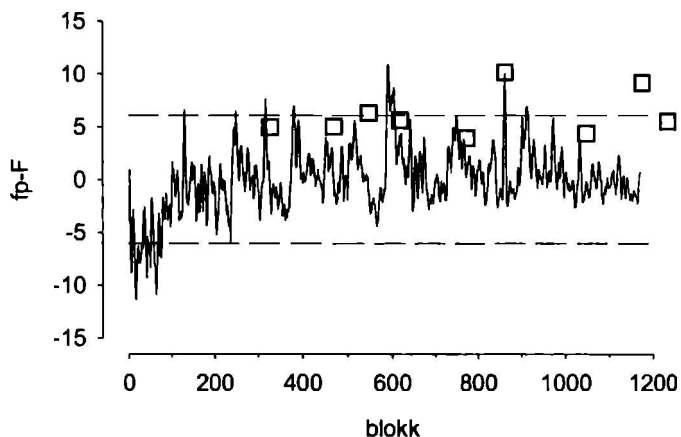
1. ábra. Kertész Imre: *Sorstalanság* című művében az újonnan bevezetett szóalakok számának változása. A folyamatos görbe az eredeti mű és a modell alapján előállított mesterséges szövegek szóalakjaiban jelentkező különbségét, a szaggatott vonal a szignifikancia-szintet mutatja. Azokat az eltéréseket tekintettük szignifikánsnak, amelyek meghaladták az $M \pm 2\sigma$ értéket.



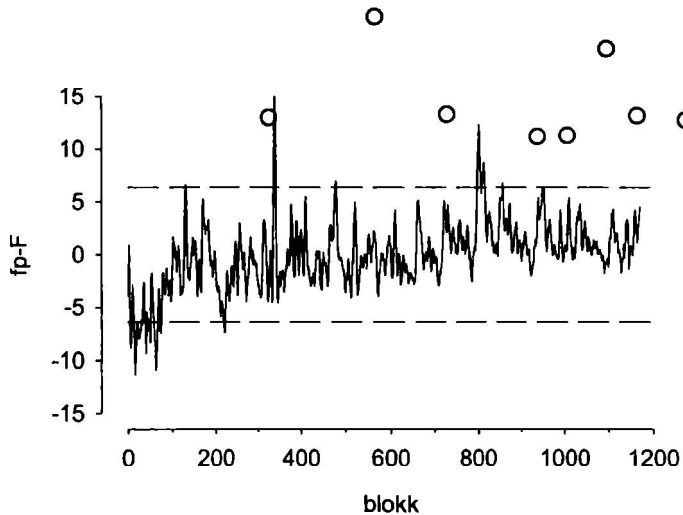
2. ábra. Kertész Imre: *Sorstalanság* című művének német fordítása: *Roman eines Schicksallosen*. A német és a magyar nyelvű szövegben, apró eltérésektől eltekintve, a szövegnek ugyanazon a pontján emelkedett meg az újonnan bevezetett szóalakok száma. A nyilak a német nyelvű szöveg azon pontjait mutatják, amelyekben az újonnan megjelenő szóalakok száma szignifikáns eltérést eredményezett.



3. ábra. Kertész Imre: *Sorstalanság* című művének angol fordítása: *Fateless*. A magyar és a német nyelvű szövegekhez hasonlóan olyan eseményeknél jelentek meg a kiugrások, amelyek nem képezik szerves részét a szövegnek, nem logikus következményei az előzményeknek, és a folytatáshoz sem kötődnek.

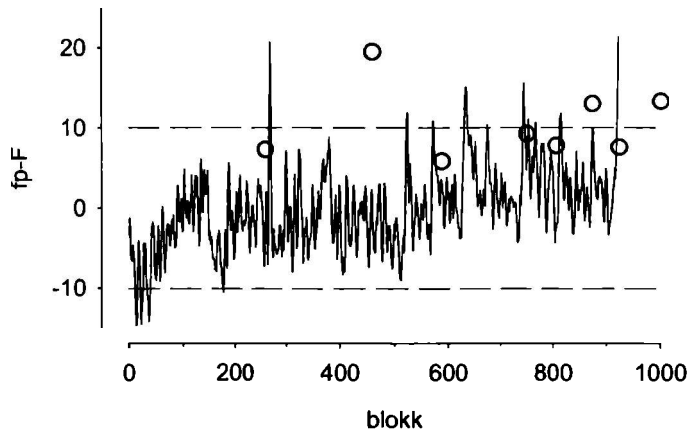


4. ábra. Kipling: *The Jungle Books* című művében (eredeti sorrend, 4. táblázat) a blokkonkénti szóalakok és a modell által számolt értékek közötti különbség. A tíz négyzet azokat a pontokat jelöli, amelyekben szignifikáns vagy közel szignifikáns eltérést mértünk az eredeti szöveg és a modell által generált mesterséges szövegek között. Az 5., 7. és a 10. pontok nem érik el a szignifikancia-szintet, de a magyar nyelvű szövegben a megfelelő szövegrészek szignifikancia-szint feletti kiugrást okoztak.

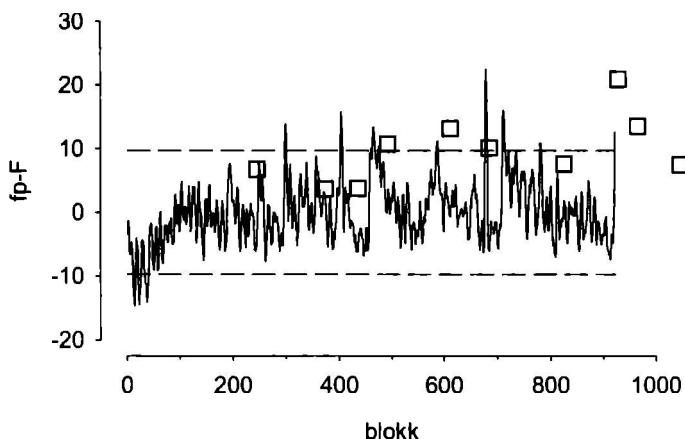


5. ábra. Kipling: *The Jungle Books* című művének egy olyan (továbbra is angol nyelvű) verziója, ahol a novellákat a magyar sorrend szerint rendeztük (3. és 5. táblázat). A magyar sorrendet figyelembe véve újra tíz olyan pontot találtunk, amelyek legalább az egyik nyelven szignifikáns eltérést okozott.

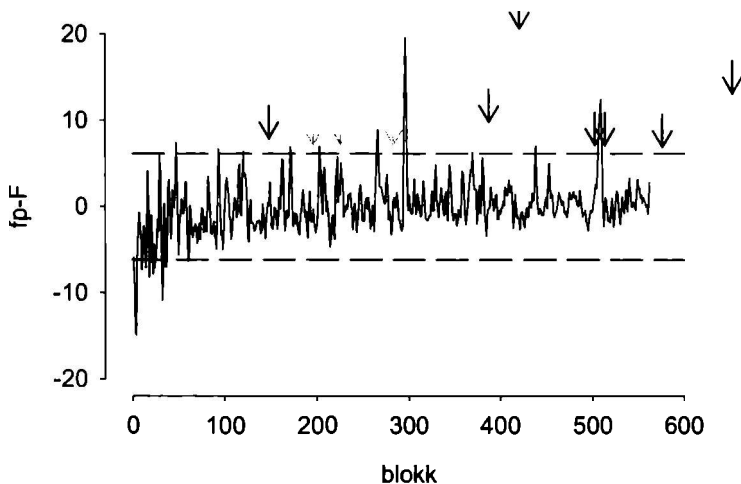
blokk



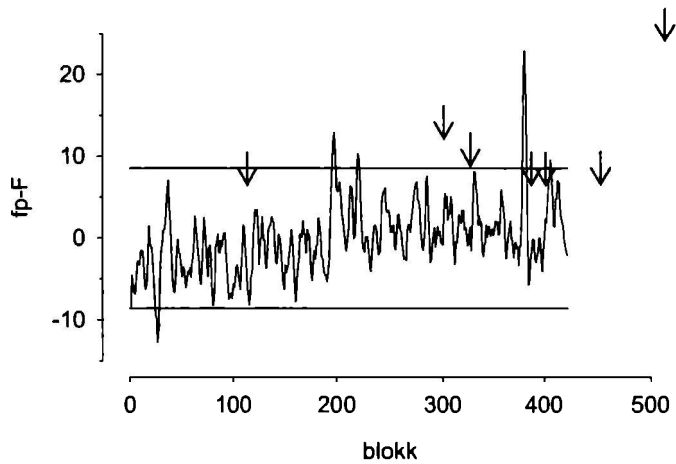
6. ábra. *A dzsungel könyve* magyar nyelvű műelemzése (6. táblázat). A magyar fordításban megváltozott a novellák eredeti sorrendje. Ez a sorrend is alkalmas arra, hogy megtaláljuk a műben azokat a helyeket, ahol megváltozott a felhasznált szókészlet; ezzel szemben az eredeti mű és fordításának közvetlen összehasonlítása a novellák eltérő sorrendje miatt nem lehetséges. Akkor tudjuk csak összehasonlítani ezeket a könyveket, ha az egyik könyvben a novellák sorrendjét hozzáigazítjuk a másik sorrendhez.



7. ábra. *A dzsungel könyve* magyar nyelvű szövege, amelyben a novellák sorrendje az eredeti, angol sorrendet követi (7. táblázat). A négyzetek, az 5. ábrához hasonlóan, a szignifikáns vagy közel szignifikáns eltéréseket mutatják. Azokat a pontokat jelöltük meg, amelyek legalább az egyik nyelven elérték a szignifikanciaszintet. Egyetlen helyen nem találunk egybeesést, ez pedig az angol nyelvű szöveg első pontja, mivel ez a magyar nyelvű szövegben nem jelent meg.



8. ábra. Az *Alice-történetek* eredeti, angol nyelvű szövegének elemzése. A fekete nyilak azokat a pontokat mutatják, ahol a magyar nyelvű szövegben is kaptunk kiugrást, míg a szürkék azokat, amelyeknél a magyar szövegben nincs kiugró szóalakszám-emelkedés.



9. ábra. Az *Alice*-történetek magyar nyelvű fordításának elemzése. A magyar nyelvű szövegben egyetlen olyan szignifikáns kiugrást kaptunk (a könyv végén egy ünnepség leírása), amely az angol szövegben nem volt jelen.

1. táblázat. Kertész Imre: *Sorstalanság* ill. a mű angol és német nyelvű fordítása; Rudyard Kipling: *The Jungle Books* és magyar fordítása; Lewis Carroll *Alice-történeteinek* eredeti, angol szövege és ezek magyar nyelvű fordításai. Az angol szövegekben fordul elő a legkevesebb különböző szóalak és ezzel párhuzamosan a legkevesebb hapax legomena.

	szövegszó	szóalak	hapax legomena
Sorstalanság	56100	14740	10253
Fateless	71600	6710	3186
Roman eines Schicksallosen	71900	9992	6043
The Jungle Books	117100	7452	3124
A dzsungel könyve	92200	20362	13372
Alice (angol)	56200	3879	1515
Alice (magyar)	42200	9730	6257

2. táblázat. Kertész Imre: *Sorstalanság* című műve és annak német és angol fordítása. A számok félkövérrel azoknak a blokkoknak a sorszámát jelölik, amelyekben az újonnan bevezetett szóalakok száma magasabb, mint az a modell alapján várható volt, tehát meghaladják az $M \pm 2\sigma$ értéket. A dőlt számok azoknak a blokkoknak a sorszáma, ahol egy másik nyelven szignifikáns eltérést találtunk, az adott szövegben pedig ahhoz közeli értéket.

	magyar	német	angol
Vili bácsi		45	43
csepeli üzem		73	71
indulás a vonattal 1			156
indulás a vonattal 2		209	
Auschwitzba érkezés	170	222	215
Buchenwaldba érkezés	262	337	329
reggeli készülődés, üzem	310	398	392
főszereplő testi állapota	354	453	447
főszereplő lelki állapota	402	518	511
kórház	429	552	
Pjetyka főz	459	587	
új transport érkezett	484		618
hazaindulás	510	651	650

3. táblázat. Kipling: *The Jungle Books* és magyar fordítása, *A dzsungel könyve* című művekben szereplő novellák a megjelenés sorrendjében.

eredeti, angol sorrend	magyar sorrend
Book I	
Mowgli's Brothers	Maugli testvérei
Kaa's Hunting	Ká vadászata
„Tiger-Tiger”	Hogyan született a félelem
The White Seal	„Tigris! Tigris!”
„Rikki-Tikki-Tavi”	A király ankusa
Toomai of the Elephants	Rátok szabadítom a dzsungelt
Servants of the Queen	A vörös kutyák
Book II	Tavaszi futás
How Fear Came	
The Miracle of Purun Bhagat	„Riki-Tiki-Tévi”
Letting in the Jungle	A fehér fóka
The Undertakers	Purun Bagát csodája
The King's Ankus	A krökodilus története
Quiquern	Kvikvern
Red Dog	Kis Tumáj és az elefántok tánca
The Spring Running	A királynő szolgálói

4. táblázat. Kipling: *The Jungle Books* című művében, megtartva a novellák eredeti, angol sorrendjét (4. ábra), meghatároztuk a mesterséges szövegek átlagától legalább 2σ -val eltérő helyeket és az ottani eseményeket.

ssz.	mese címe	esemény	blokk
1.	Kaa's Hunting	királyi palota leírása	130
2.	The White Seal	a fehér fóka leírása a több hónapos helykeresés után	249
3.	Rikki-Tikki-Tavi	Rikki-Tikki-Tavi mese kezdete	316
4.	Toomai of the Elephants	az elefánt életének leírása	381
6.	The Miracle of Purun Bhagat	barátok felsorolása foglalkozásuk szerint	591
		zarándokok felsorolása	604
8.	The King's Ankus	kincsek felsorolása	862
9.	Quiquern	Kadlu	913

5. táblázat. Kipling: *The Jungle Books* (magyar sorrend, 5. ábra) című novelláskötetében a mesterséges szövegek átlagától legalább 2σ -val eltérő helyek és az ottani események.

ssz.	cím	esemény	blokk
1.	Kaa's Hunting	királyi palota leírása	130
2.	The King's Ankus	kincsek felsorolása	339
3.	Letting in the Jungle	Hathi lerombolja a falut	479
6.	The Miracle of Purun Bhagat	barátok felsorolása foglalkozásuk szerint	801
		zarándokok felsorolása és a hegy leírása	816
7.	Undertakers	madarak felsorolása	858
8.	Quiquern	Kadlu	952

6. táblázat. Kipling: *A dzsungel könyve* című novelláskötetében (magyar sorrend, 6. ábra) a mesterséges szövegek átlagától legalább 2σ -val eltérő helyek és az ottani események.

ssz.	cím	esemény	blokk
2.	The King's Ankus	kincsek felsorolása	268
4.	Rikki-Tikki-Tavi	Rikki-Tikki-Tavi mese kezdete	523
5.	The White Seal	a fehér foka	572
6.	The Miracle of Purun Bhagat	Purun Bhagat jellemzése	633
7.	Undertakers	madarak felsorolása	674
9.	Quiquern	kezdetek	744
		Kadlu	753
		tél	767
	Quiquern + Toomai of the Elephants	záró vers + kezdő vers	814
10.	The Servants of the Queen	afganisztáni emír látogatása	873

7. táblázat. Kipling: *A dzsungel könyve* (angol sorrend, 7. ábra) című novelláskötetében a mesterséges szövegek átlagától legalább 2 σ -val eltérő helyek és az ottani események.

	cím	esemény	blokk
4.	Toomai of the Elephants	eleje	299
5.	Parade-Song of Camp Animals		404
6.	The Miracle of Purun Bhagat		466
7.	Undertakers	eleje	588
8.	The King's Ankus	kincsek	679
9.	Quiquern	eleje	711
10.		Angutivum Taina	781

8. táblázat. Lewis Caroll *Alice*-történeteinek és azok fordításainak összehasonlítása.

	angol	magyar
Hódító Vilmos száraz története	47	36
Advice from a Catterpillar vers	93	
Duchess versikéje	120	
nehézségek a flamingóval	171	
teknős tantárgyai	203	
Alice Adventure's in Wonderland végén Alice félálomban	266	198
Ykcowrebbaj és Jabberwocky (Róscaffung a és Gruffacsór) vers	296	220
Tweedledum (Subidu) verse	370-381	275-285
Humpty Dumpty (Dingidungi) vers magyarázata	438	331
Lovag éneke	509	381
Alice királynő ünneplés, ebéd		406

9. táblázat. Lewis Caroll *Alice*-történeteiből azon blokkok sorszáma, amelyekben az újonnan bevezetett szóalakok száma szignifikáns eltérést mutatott. A rövid blokkok az angol szövegben is csak éppen elérték a szignifikancia-küszöböt, míg a magyar szövegben ugyanezekben a helyeken nem vagy csak nagyon kicsi kiugrásokot találtunk ($2\sigma = 6,1486$).

blokk	új szóalak	különbség	blokk	új szóalak	különbség
47	18	7,43	295	20	13,24
			296	22	19,64
93	19	6,7	296	22	13,3
120	12	6,43	438	11	7,06
171	13	6,95	506	11	6,83
			507	12	9,63
203	12	7,03	508	14	11,12
			509	15	12,45
265	11	7,26	510	13	7,53
266	15	8,95			