

A nyelvi változás formális modellezése

1. Bevezetés

Bár a nyelv változásának elemzése és leírása mindig is a nyelvészeti érdeklődés homlokterében állt, a nyelvi változás formális modellezése viszonylag új és gyorsan fejlődő kutatási terület, amely a generatív nyelvészet, a történeti nyelvészet, a korpusznyelvészet, a nyelvelsajátítás, a populációgenetika, valamint a számítógépes és matematikai-statisztikai modellezés eszköztárait felhasználva igyekszik feltárni a nyelvek változásának mechanizmusait. Ez a fajta megközelítés felcsillantja annak a reményét, hogy a nyelvi változás pontosabb jellemzésével választ kaphatunk a történeti nyelvészet olyan klasszikus kérdéseire, mint hogy vannak-e a nyelvek változásának „természetes” irányai, a nyelvi változás visszafordíthatatlan folyamat-e; ugyanakkor pedig pontosabban meg tudjuk ragadni a nyelvelsajátítás és a nyelvi változás kapcsolatrendszerét, vagy a nyelvi változás gyorsasága és jellege, valamint a nyelvet hordozó népesség populációdinamikája közötti összefüggést.

Az alábbiakban egy rövid szakirodalmi áttekintést kívánok adni, bemutatva a jelenleg élő főbb kutatási irányzatokat. E bevezetőt követően először a nyelvváltozás-modellek általános felépítését mutatom be (2. rész), majd azok legfontosabb részeit, a nyelvmodelleket (2.1. rész), a tanulásmo-delleket (2.2. rész), a nyelvikompetencia-modelleket (2.3. rész) és a populációmodelleket (2.4. rész) tekintem át. Végül ez eddigi kutatások érdekes eredményeiből, illetve a még nyitott, vitatott kérdé-sekből ismertetek néhányat.

Fontos megjegyezni, hogy ebben az áttekintésben a (morfo)szintaktikai változásokat feldol-gozó szakirodalomra koncentrálok, a fonológiai változások formális modellezésével foglalkozó szakirodalom bemutatására nem vállalkozom.

2. A nyelvváltozás-modellek közös jellemzői

Közös jellemzője a modelleknek, hogy a nyelvi változást a nyelvelsajátítás folyamatán keresztül kívánják megragadni, követve azt az elképzelést, miszerint egy adott nyelvet beszélő két generáció nyelvváltozata közötti eltérés oka abban áll, hogy az újabb generáció a régiétől némileg eltérő nyelvtant sajátított el.

A nyelvi változás formális modelljei jellemzően a következő elemekből építkeznek:

- Nyelvmodell: milyen a nyelv felépítése, szerkezete.
- Nyelvelsajátítás-modell: milyen tanulási algoritmussal modellezhető a nyelvelsajátítás.
- Nyelvikompetencia-modell: hogyan jellemezhető egy beszélő anyanyelvi tudása.
- Populációmodell: milyen a nyelvet hordozó közösség összetétele; az egyes generációk aránya, a generációk átfedésmentesek-e, vagy sem.

A fenti modelleket természetesen úgy érdemes megkonstruálni, hogy azok minél jobban közelítsenek a valósághoz, ugyanakkor viszont formális tulajdonságaiknak köszönhetően alkalmasak legyenek a modellezésre. A fenti modulok megalkotásánál a szerzők értelemszerűen az egyes részterületek eredményeire támaszkodnak (így például a nyelvelsajátító algoritmusok meghatározásánál a nyelv-elsajátítás kutatási eredményeire).

A fenti elemeket megfelelően specifikálva jól modellezhető az, hogy egy-egy nyelvi jelenség több generáción át hogyan terjed el (vagy éppen hal ki) a nyelvet beszélő populációban. Megvizsgál-ható az is, hogy egy új nyelvi jelenség „sikere” miképpen függ az olyan tényezőktől, mint hogy a vizs-gálat kezdetén a népesség mekkora arányban használta ezt az új jelenséget, vagy hogy a nyelvelsa-játítás mechanizmusa pontosan hogyan jellemezhető.

Mivel a szövegkorpuszok segítségével sok esetben jól rekonstruálható egy-egy nyelvi változás térhódításának az időbeli lefolyása, a fenti módon konstruált modelleket tesztelhetjük úgy, hogy az általuk felrajzolt változási pályákat összevetjük a valóságban tapasztaltakkal.

Ugyanakkor a formális modellezés arra is felhasználható, hogy egyes, a történeti nyelvészeti vagy nyelvelsajátítási szakirodalomban felmerült elméleti javaslatok érvényességét, magyarázó erejét próbának vessük alá. A nyelvi változás formális modellezésével például feltárható az, hogy a nyelv-elsajátítás egyes modelljeivel a nyelvi változás milyen típusai konzisztensek.

2.1. Nyelvmodell

Az egyes modellek jelentősen eltérnek abban, hogy mennyire finoman kidolgozott nyelvmodellel operálnak. Általánosságban elmondható azonban, hogy jellemzően generatív szemléletű nyelvmo-delljük van, vagyis azt feltételezik, hogy a lehetséges nyelvek körét az elvek és paraméterek alkotta

univerzális nyelvten (UG) jelöli ki. Ez azért is előnyös, mert a nyelvészajátítás feladata ez esetben nem más, mint véges számú paraméter megfelelő értékének a megtanulása. Ez azt jelenti, hogy jelentősen korlátozva van azoknak a potenciális nyelveknek a köre, amelyek közül a megtanulandót azonosítania kell a tanulóalgoritmusnak. Ez a korlátozott úgynevezett hipotézisér (vagy keresési tér, *search space*) lehetővé teszi, hogy az modellezett beszélő a tanulóalgoritmus segítségével véges lépésben, véges számú példa alapján magas valószínűséggel eltalálja a „szülők” (a példákat generáló beszélők) nyelvtenát.

A szerzők tanulmányaikban a hangsúlyt jellemzően a tanulási algoritmusok, illetőleg a populációdinamika kérdéseire helyezik, ezért nyelvmodelljeik többnyire meglehetősen vázlatosak, és jelentősen eltérnek a kidolgozottság tekintetében. Niyogi (2002) a nyelvet gyakorlatilag a jólformáltnak minősülő szófüzerek halmazával azonosítja: a mondatokat lexikai elemek rendezett halmazának tekinti, és nem feltételez semmilyen hierarchikus szerkezetet. Yang (2002), bár tanulmányában egy szórendi változás elterjedésének meggyőző elemzését adja, a nyelvmodelljét illetően csak annyit mond ki, hogy az elvek és a paraméterek elméletének kereteit feltételezi. Briscoe (2000) tanulmánya kifinomultabb modellel dolgozik: az ő megközelítésében ugyanis a tanulóalgoritmus nemcsak a szintaktikai jólformáltságot, hanem a szófüzerek és a hozzájuk rendelt logikai forma (jelentés) megfelelőségét is figyelembe veszi. Briscoe (2002) a mondatnak egy, a kategoriális grammatikából kölcsönzött szerkezetet tulajdonít, a mondatjelentést pedig egy egzakt formális szemantika módszerrel (az úgynevezett lambda-kalkulus) segítségével, kompozicionálisan származtatja a szerkezetből és lexikális elemek jelentéséből.

2.2. Tanulásmodell

A tanulási algoritmus meghatározása a formális modell felépítésének kulcslépése. A tanulási algoritmusokat jellemzően az alábbi forrástípusokból veszik a szerzők:

- generatív történeti nyelvészeti szakirodalomban javasolt algoritmusok, például az úgynevezett küszöbinger-tanulóalgoritmus (*trigger learning algorithm*) (Lighthfoot 1992, Gibson–Wexler 1994);
- nyelvészajátítási vagy általában a tanuláspszichológiai szakirodalomban javasolt algoritmusok;
- gépi tanulás elméletében használt, elméleti egyszerűségük miatt vonzó algoritmusok.

A tanulási algoritmusok lényegében azt a folyamatot ragadják meg, és teszik explicitté, ahogyan a nyelvtenuló az újabb és újabb nyelvi adatokkal találkozáva fokozatosan módosítja a megtanulandó nyelvtenről alkotott hipotézisét (ideális esetben mindaddig, mígnem a megtanulandó nyelvvel azonos nyelvtenhez konvergál).

Niyogi (2002) cikkében három tanulóalgoritmust vizsgál. Az úgynevezett emlékezet nélküli (*memoryless learner*) és a mintaalapú hibaminimalizáló (*batched error based learner*) tanulóalgoritmusok igen egyszerűek: fő vonzerejük az, hogy egy nyelvi változás egy adott populáción belüli, generációk közötti elterjedését egzakt módon, függvényszerűen tudjuk modellezni. Ezekben a modellekben a nyelvtenuló a felnőtt generáció által generált mondatokkal találkozik, és ezek alapján dől el, hogy a felnőtt populáció által beszélt két versengő nyelvváltozat (L1 és L2) közül melyiket sajátítja el. Ezek a mondatok háromfélék lehetnek aszerint, hogy csak L1 nyelvváltozatban, csak L2 nyelvváltozatban vagy pedig mind L1 és L2 nyelvváltozatban grammatikusak. Az egyes nyelvváltozatok generációk során alakuló terjedését, illetve visszaszorulását ezért döntően befolyásolja, hogy mekkora az átfezés L1 és L2 között a grammatikus mondatok tekintetében.

Niyogi (2002) ezen kívül vizsgálja a jól ismert úgynevezett küszöbinger-tanulóalgoritmust is: eszerint egy paraméter adott értékét akkor sajátítja el a nyelvtenuló, ha az arra egyértelműen utaló mondatok egy adott küszöbértéket meghaladó arányban szerepelnek azok között a mondatok között,

amelyekkel a tanulási folyamat során találkozunk. Niyogi (2002) ezzel kapcsolatban arra az érdekes eredményre jut, hogy ha a paraméterek helyes értékének elsajátítása valóban a fenti módon történik, akkor teljesen homogén nyelvi közeg esetén is automatikusan meg fog indulni egy eltérő nyelvváltozat térhódítása. Ez szép példája annak, hogy a formális modellezés segítségével egy szakirodalmi hipotézis következményei egzaktan elemezhetők, és ezáltal a formális modellezés hozzájárulhat a szakirodalomban szereplő javaslatok értékeléséhez.

Yang (2002) cikkének homlokterében a versenyző nyelvtanok tézise áll; tanulálgörgetmus tekintetében az pszichológiai szakirodalomban ismert úgynevezett lineáris jutalmazási és büntetési sémát (*linear reward and punishment scheme*) használja.

Briscoe (2000) tanulmányában konkrét algoritmusok kiértékelése helyett azzal az általánosabb kérdéssel foglalkozik, hogy a nyelvsajátító algoritmus esetén beszélhetünk-e úgynevezett induktív aszimmetriáról (*inductive bias*), vagyis igaz-e az, hogy az algoritmus egyes nyelvváltozatokat, paraméterértékeket könnyebben elsajátít, mint másokat. Ilyen helyzet áll fenn például akkor, ha az UG egyes paraméterértékei a tanulási folyamat kezdetén nem specifikálatlanok, hanem egy úgynevezett előre beállított (*default*) értékkel vannak feltöltve. Ez esetben az algoritmus csak akkor változtatja meg ezt az értéket, ha megfelelő mennyiségű ellenpéldával szembesül a tanulási folyamat során. Ha a nyelvsajátítási folyamatban valóban működik ehhez hasonló induktív aszimmetria, az részben magyarázatot nyújthat a nyelvsajátítás mint tanulási folyamat gyorsaságára és hatékonyságára.

2.3. Nyelvikompetencia-modell

A szakirodalom egy része azt feltételezi, hogy a tanulás folyamatának lezárultával a nyelvtanuló egy konkrét nyelvtanra konvergál, és ezt követően ennek a nyelvtannak a segítségével generál nyelvi adatokat (pl. Niyogi 2002 és Briscoe 2000).

Más szerzők (pl. Yang 2002) ezzel szemben azt feltételezik, hogy a beszélők több párhuzamos nyelvtannal is rendelkezhetnek egy időben. Egy-egy nyelvtanhoz tartozik egy adott súly, és ez annak a valószínűségét fejezi ki, hogy a beszélő az adott nyelvtant használja egy nyelvi adat generálására. Ezekben a modellekben a tanulási folyamat nem más, mint az egyes nyelvtanvariánsokhoz tartozó súlyok módosítása annak függvényében, hogy a nyelvtanuló melyik nyelvtanvariáns segítségével tudja sikeresen elemezni azokat a nyelvi adatokat (mondatokat), amelyekkel a tanulás folyamán találkozunk. Ebből az is következik, hogy nem homogén környezetben (vagyis ahol a felnőtt beszélők két vagy több nyelvváltozatot beszélnek) a nyelvtanuló két vagy több nyelvtant fog elsajátítani és használni. A nyelvtanvariánsok súlya a környezet összetételétől, illetve attól függ, hogy az egyes nyelvváltozatok mennyire fedik át egymást.

2.4. Populációmodell

A populációmodellek tekintetében a szakirodalomban több, jelentősen eltérő megoldást találunk. Így például Niyogi (2002) számos leegyszerűsítő feltevéssel dolgozik: a beszélők generációi között nincs átfedés, a nyelvtanuló generáció tagjai csak a felnőtt generáció tagjaitól tanulnak (vagyis nincs generáción belüli tanulás, úgynevezett *peer learning*), az egymást követő generációk tagjai azonos valószínűséggel találkozhatnak egymással, a populáció nagysága végtelennek van tekintve stb. A fenti leegyszerűsítések előnye az, hogy így a szerző egyrészt a tanulálgörgetmusok összehasonlítására tud koncentrálni, másrészt pedig analitikusan levezethető, zárt képletbe összefoglalható, intuitíve jól átlátható és értelmezhető eredményekre jut.

Briscoe (2000) ettől jelentősen eltérő megközelítést alkalmaz: a populáció összetételét, növekedését, a populáció tagjainak interakcióját részletesen modellezi. Niyogitól (2002) eltérően megen-

gedi azt, hogy a nyelvtanulók a saját generációjuk tagjai által generált nyelvi adatból is tanuljanak, és véges létszámú populációval dolgozik: jól definiált tulajdonságokkal rendelkező ágensekből álló dinamikus rendszert vizsgál, amelyben az ágensek adott valószínűséggel, meghatározott szabályok szerint lépnek interakcióba. A részletes populációmodell valóságghűbb elemzést tesz lehetővé, ennek azonban az az ára, hogy az eredményeket nem lehet képletszerűen, zárt formában levezetni, hanem számítógépes szimulációkhoz kell folyamodni. (Cangelosi-Parisi 2002)

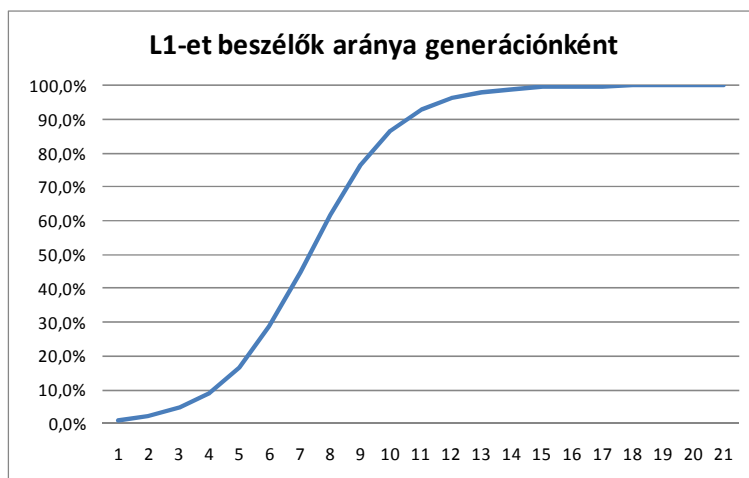
3. A nyelvi változás formális modellezésének eredményei, nyitott kérdései

A nyelvi változás formális modellezése számos alapvető összefüggés feltárásához vagy már létező intuitív elképzelések, sejtések pontosabb megragadásához és alátámasztásához segített hozzá.

A nyelvi változás gyorsaságát és irányát illetően például a szóban forgó nyelvváltozatot beszélők népességben belüli aránya, illetve a nyelvváltozatok közötti átfedés nagysága (az A nyelvtanból generált mondatok hány százaléka nem grammatikus a B nyelvtan szerint és fordítva) bizonyultak a legfontosabb paramétereknek. Mivel ezek a paraméterek korpuszokból is becsülhetők, a szakirodalomban felállított hipotézisek helytállóságát a történelmi adatok segítségével is tesztelni lehet.

Az alábbi grafikon az L1 nyelvváltozat terjedését mutatja be L2 nyelvváltozat rovására az alábbi paraméterek mellett (Niyogi 2002):

- tanulóalgoritmus: emlékezet nélküli tanuló,
- L1-et beszélők kezdeti aránya: 1%,
- annak a valószínűsége, hogy egy L1-et beszélő személy L2-ben is létező mondatot mond: 90%,
- annak a valószínűsége, hogy egy L2-et beszélő személy L1-ben is létező mondatot mond: 95%.



A nyelvi változás formális modelljei segíthetnek választ adni arra a régi dilemmára is, hogy a nyelvi változások irányukat tekintve determinisztikusak, visszafordíthatatlanok-e, illetve hogy egy részpopulációban megjelent változás milyen eséllyel terjed el a teljes populációban. Bár konkluzív válasszal még nem rendelkezünk erre a kérdésre, az már világosan látszik, hogy ebben a tekintetben sok múlik azon, hogy milyen tanulóalgoritmussal modellezzük a nyelvsajátítás folyamatát.

A nyelvi változás formális modellezése segíthet egyes, látszólag független nyelvi változások közötti összefüggések feltárásában is: így például Yang (2000) meggyőzően mutatja be két szintaktikai változás (az úgynevezett pro-drop¹ terjedése és a V2 szörend² visszaszorulása) közötti összefüggést a francia nyelvtörténetben.

Végezetül, mint már említettük, a nyelvi változás formális modellezésének egyik fő előnye az, hogy lehetővé teszi az elméleti szakirodalomban a nyelvi változás mikéntjéről fölállított hipotézisek pontos megragadását és tesztelését, értékelését.

Összességében megállapítható tehát, hogy a formális modellezés mint módszertan jelentősen hozzájárulhat a nyelvi változások jobb megértéséhez.

SZAKIRODALOM

- Briscoe, Edward J. 2000a. Grammatical Acquisition: Inductive Bias and Co-evolution of Language and the Language Acquisition Device. *Language* 76/2: 245–96.
- Briscoe, Edward J. 2000b. Evolutionary Perspectives on Diachronic Syntax. In: Susan Pintzuk – George Tsoulas – Anthony Warner (eds.): *Diachronic Syntax: Models and Mechanisms*. Oxford University Press, Oxford.
- Briscoe, Edward J. (ed.) 2002: *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*, Cambridge University Press, Cambridge.
- Clark, Robin – Roberts, Ian 1993. A Computational Model of Language Learnability and Language Change. *Linguistic Inquiry* 24/2: 299–345.
- Cangelosi, Angelo – Parisi, Domenico 2002. *Simulating the Evolution of Language*. Springer, London.
- Gibson, E. – Wexler, K. 1994. Triggers. *Linguistic Inquiry* 25: 355–407.
- Jäger, Gerhard 2007. Evolutionary Game Theory and Typology. A Case Study. *Language* 83/1: 74–109.
- Kandler, Anne – Steele, James 2008. Ecological Models of Language Competition. *Biological Theory* 3/2: 164–73.
- Kroch, Anthony 1989. Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change* 1: 199–244.
- Lightfoot, David 1992. *How to Set Parameters: Arguments from Language Change*. MIT Press, Cambridge (MA).
- Lightfoot, David 1997. Shifting Triggers and Diachronic Reanalyses. In: Ans van Kemenade – Nigel Vincent (eds.): *Parameters of Morphosyntactic Change*. Cambridge University Press, Cambridge.
- Lightfoot, David 1999. *The development of language: Acquisition, change and evolution*. Blackwell, Oxford.
- Nedergaard Thomsen, Ole 2006: *Competing Models of Linguistic Change. Evolution and beyond*. John Benjamins, Amsterdam.
- Niyogi, Partha – Berwick, Robert C. 1997. *A Dynamical Systems Model for Language Change*. In: *Complex Systems*, 11: 161–204.
- Niyogi, Partha 2002. The Computational Study of Diachronic Linguistics. In: David Lightfoot (ed.): *Syntactic Effects of Morphological Change*. Cambridge University Press, Cambridge.
- Niyogi, Partha 2006. *The Computational Nature of Language Learning and Evolution*. MIT Press, Cambridge (MA).
- Yang, Charles D. 2002. Grammar Competition and Language Change. In: David Lightfoot (ed.): *Syntactic Effects of Morphological Change*. Cambridge University Press, Cambridge.

Halm Tamás

PhD-hallgató

PPKE-BTK

Nyelvtudományi Doktori Iskola

¹ A pro-drop némileg leegyszerűsítve az a jelenség, hogy semleges mondatokban a mondat alanyát kifejező személyes névmás elhagyható, például *Futok* (magyar) vs. **Am running* (angol).

² V2 szörendűnek azokat a nyelveket tekintjük, ahol a következő alapszórendeket figyelhetjük meg: SVO, OVS, XVSO, vagyis az ige mindig a második helyen található.

SUMMARY

*Halm, Tamás***The formal modelling of language change**

The aim of this review article is to provide a short overview of the literature on the formal modelling of language change. The formal modelling of language change is a rapidly developing research programme situated at the intersection of generative linguistics, historical linguistics, corpus linguistics, language acquisition, population genetics and computational and mathematical-statistical modelling. Its research agenda includes such goals as the precise and formal characterisation of linguistic change, the analysis of the relationship between language change and language acquisition, and the exploration of the connections between the speed and character of language change on the one hand and the population dynamics of the language community on the other.

Keywords: computational linguistics, language acquisition, syntax, historical linguistics, corpus linguistics, generative linguistics