

Fonológiai jegyek felügyelet nélküli tanulása fonemikus korpuszból

Vásárhelyi Dániel

Eötvös Loránd Tudományegyetem, BTK, Elméleti Nyelvészet Program,
e-mail: vad@budling.hu

Kivonat A modern fonológiai ábrázolás központi eleme a szegmentumok megkülönböztető fonológiai *jegyekre* történő felbontása, ami lehetővé teszi a fonológiai szabályok tömörebb és plauzibilisebb megfogalmazását. Az utóbbi időben többen próbáltak érvelni ezeknek a jegyeknek és kombinatorikus viszonyaiknak, a *jegygeometriának* a veleszületett volta mellett, miközben mások a fonológiának a lexikonból való elsajátíthatósága mellett törnek lándzsát.

Az ismertető kutatás célja, hogy a konvex kombinatorikus geometriák algoritmikus jellemzésének legfrissebb eredményeit felhasználva egy memórialapú felügyelet nélküli algoritmust adjon a jegygeometria megtanulására, ezzel letéve a garast a lexikalista álláspont mellett általánosságban a nyelvi elemek és speciálisan a fonológiai szegmentumok belső struktúrájának elsajátításában.

Kulcsszavak: korpusznyelvészet, természetesnyelv-feldolgozás, jegygeometria, felügyelet nélküli tanulás

1. Bevezetés

A fonológiai jegyek eloszlásalapú elsajátítása különösen érdekes lehet annak fényében, hogy bizonyos jelenségek, mint például a szonoritásprojekció memórialapú magyarázatához a fonémák szubszegmentális ábrázolására van szükség (lásd [2]). Amennyiben a szegmentálás szintén elvégezhető kizárólag a fonológiai input alapján, akkor nincs szükség veleszületett specifikus fonológiai tudás feltételezésére.

2. Jegygeometria

A fonémák, az őket megvalósító fónok artikulációs és akusztikus tulajdonságai alapján, számos jeggyel jellemezhetők, ezek közül az egyes nyelvek választják ki, melyek kontrasztívák, azaz megkülönböztető szerepűek és melyek redundánsak.

Dresher a [3]-ban a kontraszt fonológiai szerepét vizsgálva arra a következtetésre jutott, hogy a nyelvészek a kontrasztivitásnak két egymással inkompatibilis meghatározása között ingadoztak. A teljesen specifikált minimálpárokra alapuló

és a jegyeken hierarchikus struktúrát feltételező megközelítések közül az elsőről meggyőző módon mutatja ki annak tarthatatlanságát.

A nyelvi elemek hierarchikus jegyekkel való ábrázolása az utóbbi időben a nyelvészet más területein is széles körben elterjedt.

3. Antimatroidok

A hierarchikus kapcsolatok egyik legáltalánosabb modellje a *konvex kombinatorikus geometria* vagy a vele ekvivalens *antimatroid*, egy olyan halmazrendszer, amely az alaphalmaz elemeinek egyesével való hozzáadásával (vagy elvételével) megkapható halmazokból áll.

Belátható, hogy a megkülönböztető jegyek rendszere egy antimatroidokból álló rendszert alkot, amelyben a fonémák és azok természetes osztályai mind konvex halmazok.

4. Algoritmusok

A vizsgált korpusz különféle a szerzők által interneten szabadon hozzáférhető magyar nyelvű szövegek saját algoritmussal történő fonetizálásával készült. A korpusz méretének további növelése nem okozott lényeges változást a kutatás eredményeiben.

A korpuszból először annak trigram modelljét állítottuk elő, majd azt követően minden fonémához hozzárendeltük a $_p_1p_2, p_1_p_2, p_1p_2_$ alakú környezetek egy elmosódott (fuzzy) halmazát olyan módon, hogy egy adott p fonémára a $_p_1p_2, p_1_p_2, p_1p_2_$ környezetekhez rendre a $pp_1p_2, p_1pp_2, p_1p_2p$ trigramok relatív gyakoriságát rendeltük. Rögzített 0 és 1 közötti küszöbértékre az ennél nagyobb relatív gyakoriságú környezetek halmazt alkotnak és definiálható a fonémák halmazán egy Ψ operátor olyan módon, hogy fonémák tetszőleges U halmazához hozzárendeljük azokat a fonémákat, amelyek környezethalmaza tartalmazza mindazon környezeteket, amelyeket U minden elemének környezethalmaza tartalmaz.

Amennyiben Ψ *izotón*, amelyet az a feltételezés, hogy a fonémák és a környezetek konvexek biztosít, a [4]-ben ismertetett Ψ -algoritmus egy antimatroid rendszert definiál, ami tézisünk szerint éppen a magyar fonológia jegygeometriájával azonos.

5. Eredmények

A kutatás jelenlegi szakaszában a paraméterek beállítása és a kapott antimatroid vizsgálata folyik, ami a teljes halmazrendszer mérete miatt nem egyszerű feladat, ezért a teljes halmazrendszer helyett annak kisebb fonémahalmazokra való megszorítását értékeltük.

Meglehetősen nagy ($>0,01$) küszöbértékekre a leggyakoribb fonémákra (e, a, t, n, k, l, o) megszorított rendszer meggyőzően egyezik egy lehetséges jegygeometriával, például az $\{e\}$, $\{e, a\}$, $\{e, a, o\}$ konvex halmazok megfeleltethetők egy $voc > back > round$ jegyhierarchiának.

Hivatkozások

1. Ball, Keith.: An Elementary Introduction to Modern Convex Geometry, Flavors of Geometry, MSRI Publications Volume 31, Cambridge, Massachusetts, (1997)
2. Daland, Robert, et al.: Explaining sonority projection effects, Phonology 28, Cambridge University Press, 197–234, (2011)
3. Dresher, B. Elan: The contrastive hierarchy in phonology, Toronto Working Papers in Linguistics, Vol 20, Toronto, 47–62, (2003)
4. Kempner, Yulia, et al.: Correspondance between two antimatroid algorithmic characterizations, The Electronic Journal of Combinatorics (www.combinatorics.org), Vol 10, RR44, (2003)