

Jung Janka (ELTE):

A morfológiai szegmentálás értékelése analógiás kiterjeszhetőség alapján – esettanulmány a magyar névszóragozásról

A morfológiailag komplex szavak reprezentációval és feldolgozásával kapcsolatban a kognitív tudományokban központi vitás kérdés, hogy a szóalakokat milyen egységekre bontva tároljuk és sajátítjuk el. Pszicholingvisztikai kísérletek (pl. Behrens, 2017) és elméleti nyelvészeti érvek (pl. Ackerman et al., 2009) is szólnak amellett, hogy bizonyos, egyébként több egységre szételemezhető alakok felbonthatatlan alakként való tárolása segítheti a hatékonyabb feldolgozást, az analógiás következtetést és a gyakori alakok gyorsabb elsajátítását. A kérdés az, hogy formális, számítógépes eszközökkel hogyan tudjuk mérni, hogy a nyelvsajátítás szempontjából egyes egységek egybe- vagy különlemezése hatékonyabb.

Az előadásomban egy alternatív módszert mutatok be a morfológiai szegmentálás értékelésére egy esettanulmányon keresztül, ami a magyar névszóragozás tanulhatóságáról szól. A javasolt megközelítés csak a felszíni alakokra hivatkozik és abban tér el az optimális felbontást kereső (pl. Virpioja et al., 2018), illetve a morfológiai változást szimuláló (pl. Elsner et al., 2020) információelméleti módszerektől, hogy a nyelvsajátításnak egy matematikai modelljén alapul és az alternatív felbontásokat aszerint értékeli, hogy azok mennyire segítik elő az analógiás általánosításokat.

Elméleti keretnek E. Mark Gold (1967) *Language Identification in the Limit* modelljét használom, az analógiás általánosítás formális meghatározását Angluin (1982) kizárólag pozitív példákban tanuló algoritmus alapján adom meg. A tanulás hatékonyságát az alapján értékelem, hogy mennyi a legkevesebb példa, ami a teljes részrendszer megtanulásához szükséges. A kísérletben Berwick és Pilato (1987) szintaktikai részrendszerek tanulhatóságával foglalkozó munkáját adaptálom a morfológiai mintázatok vizsgálatára és tőlük eltérően a szóalakoknak a tanulhatóság szempontjából optimális felbontását keresem. A módszert a magyar névszóragozás példáján mutatom be. Itt a birtokos alakok elemzése esetében érdekes vitás kérdés, hogy milyen egységekre bontjuk fel a szóalakokat (pl. Kiefer, 2018), ezért érdemes ezen a fragmentumon vizsgálni a szegmentálási stratégiák hatékonyságát.

A kísérlet eredményei szerint az algoritmus akkor képes hatékony általánosításra, ha bizonyos toldalékoknak (-i, -jA, -jU) nem minden előfordulását gondoljuk önálló egységnek. Ez alapján következtethetünk arra, hogy egyes, elméletileg szételemezhető névszói toldalékpárokra az analógiás kiterjeszhetőség szempontjából hasznos inkább egy egységként tekinteni. Ez összhangban van más, független érvekkel is, amik szerint ezeket a kiszegmentálható egységeket elméleti szempontból és a morfológia számítógépes kezelése szempontjából is érdemes fúziós morfémák részeként kezelni (pl. Rebrus et al., 2012). Az előadásom ennek az analógiás kiterjeszhetőségen alapuló, formális megfogalmazására mutat be egy ötletet.

Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. *Analogy in grammar: Form and acquisition*, 54-81.

Angluin, D. (1982). Inference of reversible languages. *Journal of the ACM (JACM)*, 29(3), 741-765.

Behrens, H. (2017). The role of analogy in language processing and acquisition. *The changing English language: Psycholinguistic perspectives*, 215-239.

- Berwick, R. C., & Pilato, S. (1987). Learning syntax by automata induction. *Machine learning*, 2, 9-38.
- Gold, E. M. (1967). Language identification in the limit. *Information and control*, 10(5), 447-474.
- Elsner, M., Johnson, M. B., Antetomaso, S., & Sims, A. D. (2020). Stop the morphological cycle, I want to get off: Modeling the development of fusion. *Society for Computation in Linguistics*, 3(1).
- Kiefer, F. (2018). A főnévragozás. In Kiefer (szerk): Strukturális magyar nyelvtan 3.: Morfológia (10.3 fejezet). https://mersz.hu/dokumentum/m280smny__249
- Rebrus, P., Kornai, A., & Varga, D. (2012). Egy általános célú morfológiai annotáció. *Általános Nyelvészeti Tanulmányok XXIV*, 47-80.
- Virpioja, S., Lehtonen, M., Hultén, A., Kivikari, H., Salmelin, R., & Lagus, K. (2018). Using statistical models of morphology in the search for optimal units of representation in the human mental lexicon. *Cognitive Science*, 42(3), 939-973.