

# Egy kereslet-kínálat elvű elemző működése és a koordináció kezelésének módszere



Sass Bálint

sass.balint@itk.ppke.hu

MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoport, PPKE ITK

## Bevezetés

A MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoportjában készül egy pszicholingvisztikai indíttatású, performanciaalapú, szigorúan balról-jobbra haladó magyar nyelvi elemző.

**Hogyan működik az a központi komponens, ami a konkrét balról-jobbra elemzést végzi?**

## Működés

Az elemző balról-jobbra haladva egyfajta függőségi szintaktikai elemzést végez a szövegen.

**Bemenet:** tokenek sorozata, ezeket veszi sorra az elemző

**Token:** részletes információval (szóalak, szótó, morfológia) ← balról-jobbra PoS-tagger

**Stack:** itt tároljuk sorra a tokeneket + a belőlük képzett nagyobb egységeket, frázisokat

**Működés:** szabályok alapján – jelenleg kézzel írt szabályok

**Szabály:** feltétel + eljárás:

ha az adott tokenre teljesül (egy/több) feltétel → lefut a hozzá rendelt eljárás

**Eljárás:** tipikusan az alábbiak történnek:

- 1 az adott token bekerül a stack-re
- 2 a stackelem lezárul, ha frázis végén vagyunk  
– pl.: esetrag lezár; névelő vagy ige lezárja a megelőzőt
- 3 *él* keletkezik stackelemek között: új (összetett) elem jön létre  
Ezen élek segítségével tudunk fákat (gráfokat) építeni a stacken.
- 4 az adott tokennek megfelelően valamilyen strukturális *szál* indul vagy zárul le

**Strukturális szál (röviden: szál):** két típusa van (kereslet-kínálat elv!):

- **felkínálás** – információt ad: „mi vagyok én?”
- **igény** – adott tulajdonságú elemet keres: „mi kell nekem?”

pl.: igekötőmentes ige: felkínáló szálát indít ↔ elváló igekötő: igét igénylő szálát indít  
hasonlóan: NOM-birtok; főnév-névutó . . .

A szálak típusa természetesen független az elemek sorrendjétől.

Szálak **halmaz**a: az épp futó szálakat egy halmazban tartjuk nyilván, ehhez a halmazhoz fordul az elemző minden egyes token feldolgozásakor.

A (tag)mondat végén az elemző értékelést készít, összesíti a benne lévő felsőszintű elemeket.

E reprezentáció segítségével szeretnénk megvalósítani a kitűzött célt:

a szöveg alapján válaszolni az olyan kérdésekre, hogy ki, mit csinált, hol, mikor stb.

## A koordináció kezelése

A szintaktikai elemzésben ismert nehézség a **koordináció** [1, 2].

**Alapelv:** a felsorolás/koordináció valamilyen értelemben **azonos típusú** elemekből áll.

**Megoldás:** egy speciális felkínáló szál (ún. **felsorolás-szál**) segítségével

- tároljuk (*pattern* adattag) a felsorolás kezdete óta feldolgozott felső szintű egységekről szóló információt (PoS-tag stb.)
- konjunktív elemnél (*és*, *vagy*, vessző): elindul egy felsorolás-szál  
– a konjunktív elemet megelőző egységről szóló információt is hozzávesszük
- a szál *külső* (pl. névutó, *és* utáni vessző, mondat vége)  
vagy *belső* esemény (pl. *A és A* alakú *pattern*) hatására zárul le
- lezáráskor kiértékelődik a *pattern* adattag  
nem egyszerű azonosság! – pl.: IGE.e3 + MN.NOM megengedett: **tápláló és nem romlik**
- a beazonosított felsorolásból → **új komplex elem** a stacken  
a komplex elem feje a koordinált elemek fejeinek **összessége** lesz
- az új komplex elem 1 egységként funkcionál aztán tovább,  
akár egy újabb koordinációs szerkezet egyik tagjaként

## Példák

1 A kutya megkergette és megharapta Marit . ← [1] problematikus példájának fordítása

A felsorolás lezárulását az *A és A* mintázat váltja ki.

→ alany + állítmány: *megkerget+megharap* + tárgy

2 Józsi és Pisti mellett nem fut el .

A felsorolás lezárását itt a névutó váltja ki → a létrejövő felsorolás egységhez kapcsolódik hozzá.

→ PP[mellett]: *Józsi+Pisti* + állítmány (+ alany)

3 Aláírják a finanszírozási szerződést a Budapesti Közlekedési Központ igazgatósága és a Fővárosi Közgyűlés jóváhagyásával .

Az 1 egységként azonosított terjedelmes koordinációhoz járul hozzá a birtok.

→ állítmány + tárgy + NP[-vAI] (+ alany)

3 Romulus és Remus , Róma későbbi két városalapítója egy fügefá árnyékában szopta a farkasanya tejét .

A *Romulus és Remus* koordináció lezárul a követő vessző hatására.

Majd első eleme lesz ugyanezen vessző által indított másik koordinációnak.

Végül az *A , A* mintázat miatt a *Róma későbbi két városalapítója* értelmezőként elemződik.

→ alany: *Romulus+Remus* (értelmezővel!) + NP[-bAn] + állítmány + tárgy

## Hivatkozások

[1] Sleator, D., Temperley, D.: Parsing English with a link grammar. In: Proceedings of the Third International Workshop on Parsing Technologies. (1993)

[2] Farkas, R., Vincze, V., Schmid, H.: Dependency parsing of Hungarian: Baseline results and challenges. In: Proceedings of the 13th Conference of the EACL, Avignon, France (2012) 55–65