

A szintaktikai szerkezet automatikus feltérképezése a beszédjel prozódiai elemzése alapján

Szaszák György¹, Beke András²

¹ BME Távközlési és Médiainformatikai Tanszék, Beszédakusztikai Laboratórium

² MTA Nyelvtudományi Intézet, Fonetikai Osztály

E-mail: szaszak@tmit.bme.hu; beke.andras@gmail.com

Kivonat A prozódia és a szintaktikai szerkezet közötti összefüggés aligha kérdéses, hiszen számos kutatás foglalkozott már kapcsolatukkal, illetve ezt az összefüggést számos beszédtechnológiai – elsősorban beszéd szintézisét célzó - alkalmazásban ki is használják. Az általánosan elfogadott álláspont szerint a prozódiai és a szintaktikai szerkezet szorosan összefügg ugyan, közöttük a kapcsolat azonban nem egy-egyértelműen meghatározott. Mindenesetre gyakorlati alkalmazások bizonyítják, hogy a szintaktikai elemzés alapján a prozódia jól előrejelezhető és kiválóan előállítható beszéd szintetizátor alkalmazásokban. A prozódia és a szintaxis közötti összefüggés másik irányát azonban – nevezetesen a szintaxis visszakövetettségét prozódiai jegyek alapján – eddig kevesen vizsgálták, illetve ha mégis, ezen vizsgálatok jellemzően minimál mondatpárok prozódia alapján történő elkülöníthetőségére vonatkoztak. Bár e vizsgálatok értékét nem szeretnénk alábecsülni, hiszen fontos elméleti jelentőségük van, eredményeik a gyakorlati alkalmazásokat tekintve azonban csak elvétve, nem igazán életszerű körülmények között lennének felhasználhatók. Cikkünkben ezért arra keressük a választ, hogy lehetséges-e a prozódiai szerkezet feltárása alapján szintaktikai szerkezetre vonatkozó információ kinyerése általánosabb, a mindennapi élethez jobban köthető tematika esetében. Mivel a kutatás célja az automatikus elemzhetőség vizsgálata, ezért a prozódiai szerkezet elemzését is automatikus eszközökkel valósítjuk meg. Eredményeink tanúsága szerint a beszédben a szintaktikai frázisok jelentős része jól beazonosítható, ráadásul, a szintaktikai hierarchia magasabb szintjein jól el is helyezhető. Mélyebb szinteken - többszörös beágyazásban - pontos szintaktikai szintbeli elhelyezést nem várhatunk a prozódiaától, a határok jelzése azonban megmaradhat.

Kulcsszavak: prozódia, szintaktikai elemzés, prozódiai szegmentálás, szintaktikai hierarchia, prozódiai hierarchia, szintaxis-fonológia interfész

1. Bevezetés

A prozódia és a szintaktikai szerkezet közötti összefüggést számos megközelítésben vizsgálták már, a szintaktikai és a fonológiai reprezentáció közötti interfészt

azonban eddig nem sikerült egységesen leírni. Ez nem meglepő, hiszen összetett jelenséggel állunk szemben, így az egységes modell megalkotása nem is feltétlenül volna megvalósítható elképzelés. Mindenesetre az eddigi kutatások néhány főbb ponton összecsengenek, így a szintaktikai és prozódiai szerkezetek közötti összefüggés általánosan elfogadott, természetét tekintve azonban nem teljesen feltárt. Az egyik legismertebb hipotézis Selkirk nevéhez fűződik (*prosodic structure hypothesis*), mely szerint egy-egy mondat prozódiai szerkezete nagyban - de nem teljes mértékben - függ a felszíni szintaktikai szerkezettől [11]. Más szerzők viszont amellett érvelnek, hogy a prozódiát közvetlenül és többnyire egyértelműen a szintaktikai szerkezet határozza meg [5]. A szerzők tapasztalatai alapján ez utóbbi megállapítás túlzottnak tűnik, ugyanakkor az idézett elméletnek nem térnek ki arra, hogy a prozódiai, illetve szintaktikai hierarchiában magasabban elhelyezkedő szintek sokkal biztosabban, míg a mélyebbek esetlegesebben feleltethetők meg egymásnak.

A prozódiai szerkezet az általánosan elfogadott hipotézisek szerint ([11], [4]) felülről lefelé haladva az alábbiak szerint alakul: a megnyilatkozás (utterance) *intonációs frázisokból* áll (IF), amelyek tovább bonthatók az ún. *fonológiai frázisokra* (FF). A fonológiai frázisokat pedig *fonológiai szavak* (FSz) építik fel, ezeket gyakran prozódiai szónak is hívják [11]. A hierarchia tovább finomítható egészen a szótag szintig, de a fonológiai frázisnál mélyebb egységeket a cikkben nem fogjuk használni, így a további ismertetéstől eltekintünk. A prozódiai szerkezet jól szemléltethető fával vagy a hierarchiát tükröző zárójelzéssel.

A mondatok szintaktikai elemzésekor hasonló hierarchiában gondolkodunk, amely az alapvető építőelemeket (pl. szavak) kapcsolja össze mondatokká: az egyes szavak szószerkezeteket alkotnak, ezek a szintaktikai frázisok (SzF). Az egyes frázisokba további frázisok ékelődhetnek (embedding), létrehozva a szintenként reprezentálható hierarchiát. A szintaktikai frázist általában domináns eleme (ún. fej) után nevezik el. A domináns elem az az elem, amely a frázis viselkedését az egyfel magasabb szintaktikai szinten meghatározza. Ily módon beszélhetünk névszói frázisokról (a fej névszó), igei és határozói stb. frázisokról. A szintaktikai elemzés során elterjedt a fareprezentáció.

A beszédtechnológiában az írott mondatok szintaktikai elemzése beszéd szintézis előtt elterjedt technológia [6]. Az első ilyen irányú próbálkozások egészen az 1980-as évekig nyúlnak vissza. A módszer alapja az a feltételezés, hogy a szintaktikai elemzés alapján az előállítandó beszéd prozódiai jellegzetességei igen jól előrejelezhetők. Ez tehát azt jelenti, hogy a felszíni szintaktikai szerkezet leképezhető a prozódiai szerkezetre, ráadásul a gyakorlati tapasztalatok alapján igen biztosan. Teljes leképezhetőségről azonban a beszéd szintézis esetén sem beszélhetünk, részben éppen ezzel magyarázható, hogy a beszéd szintézis alkalmazásokat miért érdemes egy-egy behatárolt tématerületre szűkíteni a minőség javítása érdekében [12].

A fordított irányú leképezés, azaz a prozódia alapján a szintaktikai viszonyokra való következtetés jóval kevésbé elterjedt, néhány – igaz, leginkább kutatási, kísérleti, de kevésbé gyakorlati – alkalmazásban azonban találkozhatunk vele. Több kutatásban is vizsgálták például egymástól jelentésben és/vagy ta-

golásban, írásjelezésben különböző, de a felépítő szavakat tekintve megegyező, ún. minimál mondatpárok elkülöníthetőségét prozódia alapján [9] (lényegét tekintve tehát jelentés-egyértelműsítés céljából). Az idézett tanulmányban Price és munkatársai következtetései alapján a prozódia alapján többségében jól elkülöníthetőek voltak a minimál párok, néhány kivételtől eltekintve. Munkájukban javaslatot is tettek olyan automatikus prozódia-címkézőre, amely normalizált időtartam adatok alapján szünetek osztályozására volt alkalmas. A prozódia alapján végzett egyértelműsítést vizsgálták már beszédfelismerésben is, leginkább itt is minimál párok elkülöníthetőségét célozva.

A beszédalapú egyértelműsítési feladatokban az előbbieken bemutatott minimál páros szemléltetés a legelterjedtebb, pedig az ily módon konstruált mondatpárok gyakran mesterkéltnek, gyakorlati alkalmazásban ritkán, de semmiképp sem univerzálisan használhatók. Ezért jelen kutatásban arra helyeztük a hangsúlyt, hogy amennyire lehetséges, általános célú és általánosan felhasználható eszközt dolgozzunk ki. Az alkalmazott megközelítés az automatikus szintaktikai és a prozódiai elemzések összevetése lesz, általános, relatíve nagy méretű beszédkorpuszon. A vizsgálat arra keresi a választ, hogy lehetséges-e a szintaktikai szerkezet legalább részleges, illetve minél teljesebb feltárása a beszédjel prozódiai elemzése alapján. Ha igen, mennyire megbízható ez az elemzés, lehetséges-e a szintaktikai hierarchia felállítása is? A kísérlethez automatikus prozódiai elemzőt használunk [13], így a lehetőségeket azonnal az automatikus elemzhetőség jelentette korlátok között értékeljük.

Cikkünk felépítése az alábbiak szerint alakul: elsőként bemutatjuk a prozódiai elemzést és a szintaktikai elemzést, a beszédkorpuszt. Ezt követi a kísérleti feltételek részletes leírása, a kiértékeléshez használt mérőszámok bemutatása, az eredmények ismertetése és a következtetések származtatása.

2. Beszédjel automatikus prozódiai szegmentálása

A prozódiai szerkezet feltérképezésére a beszédjelen *prozódiai szegmentálást* végzünk. Az eljárást részletesen bemutattuk már [14], [13], így itt csak a lényegesebb jellemzőit foglaljuk össze. A prozódiai szegmentáló feladata fonológiai frázisok (FF) illesztése a beszédjelhez. Ehhez a szegmentáló 7 beépített fonológiai frázismodellt tárol rejtett Markov-modell formájában (lásd 1. táblázat). Az illesztés a hangsúlyok és a dallammenetek együttes figyelembevételével történik. A felhasznált akusztikai jellemzők az alaphérfrekvencia- és az energiamenet, kinyerésüket a következő, 2.1 alfejezetben röviden áttekintjük. A fonológiai frázisokra úgy tekintünk, mint a legkisebb, önálló hangsúllyal és dallammenettel jellemezhető egységre [4]. A magyar nyelvben kijelentő módban a tipikus FF elején a hangsúlynak megfelelő kiemelést tapasztalunk, amelyet lassan ereszkedő dallammenet követ a következő hangsúlyos egységig. Ezt tekintjük a FF prototípusának (*fs*). Mivel azonban a fonológiai frázisok intonációs frázisokba, illetve megnyilatkozás-egységekbe - olvasott beszédben mondatokba, spontán beszédben virtuális mondatokba - szerveződnek, magasabb szintű tényezők is befolyásolják a hangsúlyozást és a dallammenetek alakulását. Emiatt az osztályozáshoz/illesztéshez

további FF-ok elkülönítése szükséges: a tagmondat eleje (*me*) és a tagmondat vége (*mv*) jellemzően befolyásolja a FF prototípusát, akárcsak a fókusz (*fe*) és a folytatást jelző dallammenet-emelkedés (*fv*). Ez utóbbi a következő fonológiai frázist olykor inverz hangsúlyba fordítja, azaz kiemelkedés helyett a prozódiai jellemzők lokális minimumot adnak (*s*). A prozódiai szegmentáló kimenetén tehát az illesztett fonológiai frázisok jelennek meg kezdő- és végidőpontjaikkal.

1. táblázat. A prozódiai szegmentáláshoz modellezett fonológiai frázistípusok.

Címke	FF típus
me	Tagmondat eleje
fe	Erős hangsúly
fs	Prototípus
mv	Tagmondat vége
fv	Folytatást jelző
s	Inverz hangsúly
sil	Csend

A prozódiai szegmentálás során a fonológiai frázisok egymáshoz kapcsolódási szabályszerűségeit leíró, prozódiai-nyelvi jellegű modellt is használunk. Ez a modell teszi lehetővé egyrészt az illesztést (milyen FF milyen FF után milyen valószínűséggel következhet), másrészt előkészíti a szintaktikai szerkezetre való leképezést, hiszen a prozódiai szegmentáló FF-modelljei a mondatokban, tagmondatokban elfoglalt helyük, szerepük szerint lettek kialakítva. A használt modell éppen a mondatok, virtuális mondatok (idealizált) felépítését adja meg: minden mondat tagmondat eleje frázissal (*me*) indít és tagmondat vége frázissal (*mv*) zár. Közben erősen (*fe*) és közepesen hangsúlyos (*fs*, prototípus) fonológiai frázisok tetszőleges sorrendben váltakoznak, esetleges folytatást jelző frázisokkal (*fv*). Ez utóbbit tagmondat eleje frázis (*me*) vagy inverz hangsúlyt tartalmazó frázis (*s*) követheti. Kivételes esetben mondat vége is lehet (pl. kérdés esetén). A mondatok között szünetet feltételezünk (*sil*). Fontosnak tartjuk megvilágítani, hogy az alkalmazott illesztési eljárás nem pusztán egyes prozódiaeseményhez köthető jelölők (pl. szünetjelölők, hangsúlyjelölők) detektálásán alapul (vö. ToBI, [12]), hanem a prozódiai, illetve a hozzá társított akusztikai jellemzők folyamatos követését biztosítja, így módon véleményünk szerint rugalmasabb és egységesebb prozódiai szegmentálást tesz lehetővé, lényegében az egyes detektálandó eseményeket a fonológiai frázisok modelljei inkorporálják.

2.1. Akusztikai-prozódiai előfeldolgozás

Az akusztikai-prozódiai előfeldolgozás a [13] irodalomban ismertettek alapján történik, de az egyes jellemzők kinyerésénél használt konstansok értékeit az alábbiak szerint állítottuk be: az alapfrekvencia (F_0) kinyerése ESPS algoritmussal történik 25 *ms* hosszúságú, csúsztatott ablakolással. Az energia kinyeréséhez használt ablak is 25 *ms*. A keretidő mindkét jellemzőre 10 *ms*. A nyert

alapfrekvencia-menetet ezután oktávugrásoktól szűrjük, majd 5 pontos átlagoló szűrővel simítjuk. Ezután az alapfrekvenciát logaritmikus tartományban lineárisan extrapoláljuk a zöngétlen helyeken, de csak akkor, ha a zöngétlen szakasz nem hosszabb 150 *ms*-nál és ha a zöngétlen szakasz után az alapfrekvencia nem indul túl magasról (nem emelkedhet többet 10%-nál a zöngétlen szakasz előttihez képest. Erre azért van szükség, hogy a frázisok közötti, levegővétellel nem társuló szünetet ne hogy zöngétlen beszédhangszakasznak vegyük. Az így előfeldolgozott jellemzőkhöz delta és delta-delta együtthatókat fűztünk. Az előfeldolgozás minden egyéb tekintetben azonos a [13] irodalomban bemutatottal.

2.2. Prozódiai szegmentálás és szóhatárok detektálása

Korábbi munkákban [13] [14] [2] vizsgáltuk már a szóhatárok detektálhatóságát prozódiai jellemzők segítségével. Ennek egyik útja szintén a fonológiai frázisok illesztése volt, amely magyar nyelvre a kötött hangsúlyozást kihasználva szóhatárok detektálását tette lehetővé, 77% körüli pontossággal és 57% körüli hatékonysággal magyar nyelvre, 69% körüli pontossággal és 76% körüli hatékonysággal pedig finn nyelvre. A szóhatár-detektálás vizsgálatokor nem végeztünk szintaktikai elemzést, viszont hipotézisünk, hogy a szintaktikai frázisok határa eső szóhatárok jobban detektálhatók, mint a frázisok belsejébe esők (igaz, a szóhatárok jelentős részén szintaktikai frázis határa is van). A szóhatár-detektálás elsősorban a gépi beszédfelismerést segítette, míg a szintaktikai elemzés – ha lehetséges a prozódia alapján – a beszéd gépi elemzését teheti lehetővé, amely kiemelt fontosságú az átfogóbb, gépi beszédértést/-elemzést is igénylő rendszerekben (pl. gépi tolmácsolás).

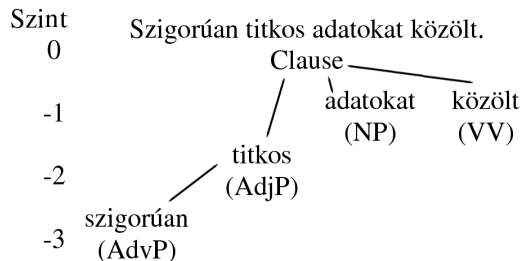
3. Szintaktikai elemzés (szövegalapú)

A szintaktikai elemzéshez a szabadon elérhető HunPars eszköz állt rendelkezésünkre [1]. Ez a szintaktikai elemző belső erőforrásokként ún. frázisstruktúra nyelvtant és lexikai adatbázisokat és a HunMorph morfológiai elemzőt használja fel. A morfológiai elemző használata nagyban emeli a komplexitást, de a magyar nyelv sajátosságai miatt aligha megkerülhető. A szintaktikai elemző kimenetén az elemzett mondat tagekkel ellátva és a szintaktikai hierarchiában elfoglalt helyzetet tükröző zárójelzéssel jelenik meg, amelyből fastruktúrájú reprezentáció is generálható. Az elemző minden lehetséges hipotézist megad, ez hosszabb mondatokra több tíz, kirívó esetben néhány száz lehetséges elemzés is lehet. Miután a prozodiát ezúttal nem egyértelműsítésre kívánjuk felhasználni, az elemzéseket szakértő kézzel egyértelműsítette. Az egyértelműsítés egyébként leginkább egyes lexikai elemek több jelentéséből adódóan vált szükségessé (pl. az 'egy' szót mindig háromféle elemzésben kapjuk meg (határozó, névelő, számnév), ha több nem egyértelmű elem is van a mondatban, akkor a hipotézisek száma összeszorozódik).

4. Anyag és módszer

A kísérleti anyagot a BABEL magyar nyelvű adatbázis [10] szolgáltatta, amely 5-7 mondatból álló bekezdéseket is tartalmaz. Ebből 330 mondatot elemeztünk (az ismétlődések miatt 155 különböző mondatot kellett csak szintaktikailag elemezni) 60 beszélőtől (30 férfi, 30 nő). Elsőként a 155 különböző mondat szintaktikai elemzését végeztük el. Ezután az egyes felvételeket beszédhang szinten szegmentáltuk kényszerített illesztéssel. A beszédhang szintű szegmentálásból kinyertük az egyes szintaktikai egységek határához köthető időpontokat. Ezeket fogjuk a prozódiai szegmentálás eredményeként előálló fonológiai frázisok határaival összevetni. Az összevetést szintaktikai szintenként végezzük elkülönítve, mivel a hipotézisünk az, hogy a magasabb szintaktikai szinteket a prozódia jobban tükrözi. Az elkülönített szintaktikai szinteket számmal jelöltük, felülről lefelé haladva: 0, -1, -2, -3, -4 (vö. 1. ábra). A mondatokat tagmondatokra bontjuk, így kapjuk a 0. szintet. A tagmondatokat szintaktikailag tovább elemezve egymásba ágyazott szintaktikai frázisokat találunk, ezeket reprezentálják a negatív számmal jelölt szintek. Míg a beágyazást nem tartalmazó szintaktikai frázisok (-1. szint) és az egyszeres beágyazást tartalmazók (-2. szint, a legtöbb jelzős szerkezet ilyen) igen gyakoriak, addig kétszeres vagy többszörös beágyazás már viszonylag ritka (lásd a 2. és 3. táblázatokban).

A szintaktikai és a fonológiai frázishatárokat akkor tekintettük egybeesőnek, ha közöttük kezdőidőpontjukat (végidőpontjukat) tekintve 150 ms-ot meghaladó időbeli eltérés nem volt. Ezt a tűrésintervallumot az alábbi megfontolások alapján jelöltük ki: (i) az intervallumnak lehetővé kell tennie kb. fél szótag nagyságrendű eltérést, mert a prozódiai szegmentáló pontossága ilyen nagyságrendű (vö. [13]), illetve (ii) mert a referenciaként vizsgált szintaktikai határokat automatikus szegmentálással határoztuk meg, ami pontatlanabb szegmentálást jelent a kézzel végzettnél. A választott tűréshatáron belül így még biztosított, hogy (iii) a prozódiai szegmentáló által illesztett fonológiai frázisok várható hossza jóval nagyobb 150 ms-nál (a vizsgált korpuszon az átlagos FF-hossz 618 ms, 211 ms szórás mellett). A fonológiai frázisok kezdetét mindig a szintaktikai frázisok kezdetével, a FF-ok végeit mindig a SzF-ok végével vetettük össze.



1. ábra. Szintaktikai szintek hierarchikus reprezentációjában

5. Eredmények és értékelés

5.1. Szintaktikai frázisok behatárolása

Az első kísérlet arra irányult, hogy megvizsgáljuk, a szintaktikai frázisok mennyire határolhatók be a prozódia alapján. Mérőszámként a *recall* értéket használjuk, mely definíció szerint:

$$Recall = \frac{tp}{tp + fn}, \quad (1)$$

itt tp a helyesen azonosított szintaktikai határok száma (true positives), fn pedig a nem azonosított szintaktikai határok száma (false negatives). Az eredmények a 2. táblázatban láthatók, külön a frázisok kezdetére és a végére. Már említettük, hogy a kiértékelést szintaktikai szintekre bontva külön-külön végezzük. Egyes esetekben (nem is ritkán) több, különböző szintű szintaktikai határt találunk egy helyen (pl. az "ügyes ember" frázisban egyszerre indul a $-1.$ és a $-2.$ szintű szintaktikai frázis, a $-2.$ szintű az "ügyes", az $-1.$ szintű az "ember" után ér véget). A kiértékelést emiatt két szára bontottuk: az egyik szálon egy helyen egyetlen, a legmagasabb szintű szintaktikai határt számoljuk csak (erre az $1B/W$ jelöléssel utalunk), míg a másik szálon az egy helyen található valamennyi szintaktikai határt egyszerre figyelembe vesszük (tehát utóbbi esetben helyes detektálás esetén valamennyi szinten egy helyes detektálást számítunk, holott "több legyet ütöttünk egy csapásra". Ugyanakkor ha elvétjük a határt, akkor azt természetesen valamennyi szinten hibaként vesszük figyelembe. Erre a számítási módra a MB/W jelöléssel utalunk).

2. táblázat. Szintaktikai frázisok határainak detektálása (recall). $1B/L$ = egy (a legmagasabb szintű) szintaktikai határ egy helyen; MB/W = több szintaktikai határ is lehet egy helyen.

Szintaktikai szint	Kezdet		Vég		Előf. száma (MB/W)
	1B/W	MB/W	1B/W	MB/W	
0	0,85	0,85	0,79	0,79	3124
-1	0,45	0,70	0,48	0,68	10339
-2	0,42	0,70	0,48	0,69	5763
-3	0,44	0,74	0,45	0,65	814
-4	0,48	0,70	0,50	0,67	187
Összes szint	0,54	0,72	0,55	0,69	20227

Az átlagos recall érték 71% (MB/W), illetve 55% ($1B/W$), amely a tagmondatok szintjén jelentősen magasabb: 85% (fráziskezdet) és 79% (frázisvég). Az eredmények statisztikai alátámasztására Kruskal-Wallis próbát végeztünk, amely igazolta, hogy a fonológiai és a szintaktikai frázisok között szignifikáns összefüggés van ($\chi^2 = 6430,606; p < 0,000$).

A megfelelő SzF kezdő- és végidőpontokat párba állítva és a recall értékeit vizsgálva Mann-Whitney és Wilcoxon W tesztekkel a tagmondatok esetén a tagmondat kezdetét szignifikánsan jobban lehet detektálni, mint a végét ($Z = -7,807; p < 0,000$). Mélyebb szintaktikai szinteken azonban megszűnik a szignifikáns különbség a kezdő és végidőpontok tekintetében (–1. szint: $Z = -0,407; p > 0,1$; –2. szint: $Z = -0,016; p > 0,1$; hasonlóan a mélyebb szintekre is).

A tagmondat szintnél mélyebb szinteken a recall értékek szinte azonosak, ebből arra következtethetünk, hogy a prozódia a szintaktikai hierarchiában elfoglalt helyzettől függetlenül jelez szintaktikai frázishatár-információt: nincs szignifikáns különbség a recall értékek között a szintaktikai szint függvényében a tagmondatnál mélyebben: ($\chi^2 = 0,224; p > 0,1$). Tehát minden SzF önálló entitásként viselkedik, függetlenül a szintaktikai hierarchiában elfoglalt helyétől.

5.2. Szintaktikai szintek elkülönítése a prozódia alapján

A következő lépésben azt vizsgáltuk, mennyire különíthetők el az egyes szintaktikai szintek a fonológiai frázisokra történő szegmentálás alapján, illetve van-e olyan FF, amely valamely szintaktikai szinthez társítható (a frázistípusok elkülönítésénél használt metodika alapján hipotézisünk, hogy kell lennie). Ha a FF típusa alapján különbséget tudunk tenni a szintaktikai szintek között, az nagyban emelné a prozódiai szegmentálás értékét az elemzésben. Azt is jó lenne tudnunk, mennyire megbízható a detektálás az egyes fonológiai frázisok típusától függően (ha van közöttük különbség). A választott mértékünk a precision:

$$Precision = \frac{tp}{tp + fp}, \quad (2)$$

ahol tp ismét a FF-ok által helyesen (150 ms-on belül) jelzett SzF határ, míg fp a beszúrt FF határok száma (amelyek tehát nem esnek egybe SzF-sal). A precision mérőszám mellett specificitás jelleggel azt is vizsgáljuk, hogy fonológiai frázistípusokra bontva hogyan alakulnak a szintenkénti relatív gyakoriságok (milyen típusú FF milyen szintű SzF-nak felel meg leggyakrabban/tipikusan). Az eredményeket a 3. és 4. táblázatokban mutatjuk be, külön frázisok elejének és végének összehasonlítására. A relatív gyakoriságok mellett az utolsó oszlopban a FF-hoz tartozó precision értéke is megtalálható.

A 3. táblázat eredményei szerint a *me* FF 86% relatív gyakorisággal tagmondat kezdetét jelöli. A –1. szintű szintaktikai frázis kezdetére a *fe*, *fs*, *mv*, illetve kisebb mértékben a *fv* fonológiai frázisok utalnak. Az *s* típusú frázis kezdete nem egyértelmű szintaktikai utalás szempontjából. A –2. szintaktikai szinttől mélyebben a FF-ok eloszlása lényegében egyenletes az egyes szintek között, így a FF típusa nem utal a szintaktikai szintre. Az eredmények összességében tehát azt jelentik, hogy a tagmondatok kezdete igen biztosan előrejelezhető a FF típusa alapján (0. szint), illetve hogy a –1. szint ettől és a mélyebben fekvő szintektől még jól elkülöníthető. Tehát a szintaktikai hierarchia prozódiai szemszögből 3 szintre tagolódik, a 0. szintaktikai szintre, a –1. szintaktikai szintre és

3. táblázat. SzF szintek és FF-ok típusának kapcsolata frázisok elején (relatív gyakoriságok) és precision.

FF típusa	Szintaktikai szint				Előfordulások száma (összes)	Precision
	0	-1	-2	-3		
me	0,86	0,07	0,04	0,02	1736	0,84
fe	0,12	0,78	0,07	0,02	2517	0,58
fs	0,09	0,83	0,06	0,01	1399	0,55
mv	0,14	0,80	0,04	0,02	2094	0,46
fv	0,22	0,72	0,04	0,01	1326	0,51
s	0,50	0,41	0,07	0,02	1456	0,57
Összes FF	0,36	0,56	0,05	0,02	10539	0,58

az összevont $-2.$ – $N.$ mélyebb szintekre. Arra is tekintettel, hogy a szintaktikai hierarchiában a mélyebb szintek felé haladva a SzF előfordulások gyakorisága radikálisan csökken, tehát igen ritkák a kettőnél többször beagyazott frázisok (vö. 2. táblázat), a fonológiai frázis segítségével behatárolt szintaktikai frázisok jelentős hányadáról tehát eldönthető, hogy nagy valószínűséggel milyen szinthez tartoznak. Az összes $-2.$ szintű és mélyebben elhelyezkedő frázis valójában több mint 85%-ban $-2.$ szintű frázisnak felel meg, csak a fennmaradó szűk 15% az ennél mélyebb szinten elhelyezkedő. Közöttük viszont a prozódia alapján különbséget nem tudtunk tenni.

4. táblázat. SzF szintek és FF-ok típusának kapcsolata frázisok végén (relatív gyakoriságok) és precision.

FF típusa	Szintaktikai szint				Előfordulások száma (összes)	Precision
	0	-1	-2	-3		
me	0,05	0,74	0,11	0,08	1736	0,58
fe	0,09	0,68	0,20	0,03	2517	0,64
fs	0,08	0,68	0,18	0,04	1399	0,60
mv	0,83	0,11	0,04	0,02	2094	0,80
fv	0,60	0,28	0,09	0,03	1326	0,73
s	0,13	0,64	0,17	0,06	1467	0,57
Összes típus	0,34	0,49	0,13	0,04	10593	0,66

A 4. táblázat eredményei szerint a frázisok végét vizsgálva a detektált *mv* típusú FF 83% relatív gyakorisággal jelezte a 0. szintű tagmondat végét. Az *fv* típusú FF gyakran (60%) szintén tagmondat végét jelzi (0. szint), azonban viszonylag gyakran jelezheti $-1.$ szintű szintaktikai frázis végét is (28%). Az *me* típusú FF vége egyértelműbben a $-1.$ szinthez kapcsolható 74% gyakorisággal, míg az *fe*, *fs* és *s* típusú FF-ok vége $-1.$ vagy $-2.$ szinten jelzi a SzF-ok végét. Ellentétben a frázisok elejére végzett vizsgálatokkal, a frázisok végét vizsgálva

már a -1 . és a -2 , illetve mélyebb szintek sem különíthetők el az illesztett FF típusa alapján a relatív gyakoriságok vizsgálatával. Ehhez tehát a frázisok elejét kell vizsgálnunk. A gyakorlatban természetesen a frázisok elejét és végét együttesen tudjuk vizsgálni az esetek döntő többségében, hiszen a frázisok végén rendszerint újabb frázisok kezdődnek (kivéve a megnyilatkozás végén és hosszabb csend előtt, bár ez utóbbi szintén informatív elem, hiszen előtte – legalábbis olvasott beszédben – a szintaktikai frázis, sőt a tagmondat is általában lezárt.

A precision és recall mérőszámok értékeit redukált FF elemhalmazzal is számítottuk annak vizsgálatára, hogy ily módon esetleg egyértelműbben lehetséges-e a szintaktikai szintek elkülönítése. A redukált FF halmazzal történő vizsgálat során a prozódiai szegmentáló nem illesztheti az fs és az s FF-okat. Utóbbit azért zárjuk ki, mert a frázisok elejére végzett vizsgálatkor nem jelezte egyértelműen a szintaktikai szintet, előbbit pedig azért, mert szerepét várhatóan az erősebben hangsúlyos, de dallammenetben nem különböző fe típusú FF modellje részben átveheti. A redukált FF elemhalmazzal végzett vizsgálatok eredményei a frázisok elejét vizsgálva az 5. táblázatban láthatók. A recall értéke visszaesik (átlagosan 48%-ra, 1B/W esetben), tehát a redukált elemhalmazzal kevesebb szintaktikai frázis kezdetét tudjuk meghatározni, ugyanakkor a precision értéke szignifikánsan nem változik. Ami miatt mégis érdemes lehet a vizsgálatot elvégezni, hogy a 0., tagmondat szintet sokkal biztosabban kiemeli. A frázisok végét vizsgálva hasonló eredményeket kaptunk: gyengébb recall mellett szignifikánsan nem jobb precision, a 0. és a -1 . szintek elkülöníthetősége javul, a -2 . szintet pedig érdemben már nem detektálja a rendszer.

5. táblázat. SzF szintek és FF-ok típusának kapcsolata frázisok elején redukált FF elemhalmazzal (relatív gyakoriságok); precision és 1B/W recall az egyes szintaktikai szintekre.

FF típusa	Szintaktikai szint				Előfordulások száma (összes)	Precision
	0	-1	-2	-3		
me	0,88	0,07	0,02	0,02	1835	0,92
fe	0,13	0,77	0,07	0,02	3455	0,58
mv	0,26	0,67	0,04	0,02	1914	0,53
fv	0,37	0,58	0,04	0,01	1782	0,57
Összes típus	0,42	0,51	0,05	0,02	8986	0,64
Recall	0,80	0,39	0,34	0,37	Átl. recall: 0,48	

5.3. Összefüggés a fonológiai és a szintaktikai frázis típusa között

Végezetül azt is vizsgáltuk, hogy felfedezhető-e valamiféle összefüggés a fonológiai frázis típusa (me, fe, fs, mv, fv, s), illetve a szintaktikai frázis típusa között (NP, AdjP, AdvP, NumP, VV, VV-Inf, PostpP). Az eredmények tanúsága szerint ilyen összefüggés a magyar nyelvben nem mutatható ki ($\chi^2 = 0,349; p > 0,1$),

a fonológiai frázisok véletlenszerűen kombinálódnak a szintaktikai frázisokkal. A frázistípusok össze nem függése a magyar nyelvben a kötetlen szórend miatt nem meglepő, a vizsgálatot érdemes lenne más, a szemantikai összefüggéseket szórenddel érzékeltető nyelven is elvégezni.

6. Összefoglalás és kitekintés

Cikkünkben a szintaktikai szerkezet feltérképezhetőségét vizsgáltuk olvasott beszédben. Egy prozódiai szegmentáló kimenete alapján a szintaktikai frázisok határait azonosítottuk, és vizsgáltuk a szintaktikai hierarchiához rendelt szintek visszakövethetőségét is pusztán a beszédjel prozódiaja alapján. A tagmondat-határok akár 92%-a, a tagmondatban elhelyezkedő, akár egymásba is ágyazott szintaktikai frázisok határainak 50-70%-a volt automatikusan meghatározható. A tagmondathatárok detektálásában a pontosságot jellemző precision mérőszám maximális értéke 84% volt, a beágyazott szintaktikai frázisokra 46 és 58% között alakult. Végkövetkeztetéseink az alábbiak: a prozódia olvasott beszédben (i) a szintaktikai határokat jól jelzi, (ii) többnyire világosan elkülöníti a tagmondathatárokat a szószerkezetek határaitól, (iii) a FF-ok/SzF-ok elejét összevetve az egyszeres beágyazódások még esetenként megkülönböztethetők (-1. és -2. szintek elkülönítése), a mélyebb szintaktikai szintek viszont egybeolvadnak, határaik azonban esetenként továbbra is detektálhatók. Ezek alapján a prozódia ütemező, szinkronizáló szerepe feltételezhető a humán beszédpercepcióban, amelyet szerényebb rétegző szerep egészít ki (0., -1. és -2. és mélyebb szintek elkülönítése).

A prozódiai és szintaktikai szerkezet összefüggéseit spontán beszédben is vizsgáljuk, ezek a kísérletek azonban még folyamatban vannak – reményeink szerint előadásunkban már az eredményekből is ízelítőt adhatunk. Spontán beszéd esetében a prozódiai szegmentálás nagyjából elvégezhető, ugyanakkor számolni kell az elemzést megnehezítő elemek megjelenésével: érzelmi töltet, amely a prozódiaát is befolyásolja; nagyobb dinamikataromány (ez az előfeldolgozásban - oktávugrás elleni szűrésben és interpolálásban - okozhat nehézségeket; a hangsúlyozási-hanglejtési "szokásjog" gyakori megszegése, dinamikus változása). A spontán beszéd szintaktikai elemzése igen nehéz feladatnak bizonyul, mivel nem tartalmaz jól körülhatárolható, egyértelműen meghatározható mondatokat. Áthidaló megoldásként ún. virtuális mondatok elemzését fogjuk elvégezni (ez alatt a spontán beszédbeli megnyilatkozások olvasott beszédhez hasonló mondatszerű formára konvertált alakját értjük - vö. [3], [7]). Továbbra is problémát jelentenek azonban a megakadásjelenségek, befejezetlen gondolatok stb., amelyek a prozódiai és virtualizált szintaktikai szerkezet egymásra képezését jelentősen nehezíthetik.

Köszönetnyilvánítás

A szerzők köszönetüket fejezik ki Nagy Katalinnak, a BME villamosmérnök hallgatójának a bemutatott munkában nyújtott segítségéért.

Hivatkozások

1. Babarczy A., Bálint G., Hamp G., Kárpáti A., Rung A., Szakadát I.: Hunpars: mondattani elemző alkalmazás, III. Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Magyarország, 2005. pp. 20-28.
2. Beke András, Szaszák György: Szótagok automatikus osztályozása spontán beszédben spektrális és prozódiai jellemzők alapján, VII. Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Magyarország, 2010. pp. 236-248.
3. Gósy Mária: Virtuális mondatok a spontán beszédben, Beszédkutatás 2003, MTA Nyelvtudományi Intézet, Budapest, 2003. pp. 19-43.
4. Hunyadi László: Hungarian Sentence Prosody and Universal Grammar, Peter Lang, 2002.
5. Kaisse, Ellen M.: Connected Speech: The Interaction of Syntax and Phonology, Academic Press, San Diego, 1985.
6. Koutny Iлона: Parsing Hungarian Sentences in order to Determine their Prosodic Structures in a Multilingual TTS system, Proc. of the Eurospeech'99 International Conference on Speech Communication and Technology, pp. 2091-2094, Budapest, Hungary, 1999.
7. Markó Alexandra: A spontán beszéd néhány szupraszegmentális jellegzetessége: Monologikus és dialogikus szövegek összevetése, valamint a hümmögés vizsgálata, PhD értekezés, ELTE, Budapest, 2005.
8. Olaszgy Gábor, Németh Géza, Olaszgy Péter: Automatic Prosody Generation - a Model for Hungarian, In: European Conference on Speech Communication and Technology (Eurospeech 2001). Aalborg, Dánia, 2001. pp. 525-528.
9. Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S., Fong, C.: The use of prosody for syntactic disambiguation, Journal of the Acoustical Society of America 90(6):2956-2970, 1991.
10. Roach, P. et al.: BABEL: An Eastern European multi-language database, Proc. of the 4th International Conference on Speech and Language Processing, Philadelphia, USA, Vol 3. pp. 1892-1893, 1996.
11. Selkirk, Elisabeth: The Syntax-Phonology Interface, in Smelser, N.J. and Baltes, Paul B. [Eds], International Encyclopaedia of the Social and Behavioural Sciences, 15407-15412, Oxford: Pergamon, 2001.
12. Silverman, K.: On costumizing prosody in speech synthesis: names and addresses as a case in point, in Proc. ARPA Workshop on Human Language Technology, pp. 317-322, 1993.
13. Szaszák György: A szupraszegmentális jellemzők szerepe és felhasználása a gépi beszéd felismerésben, PhD értekezés. Budapesti Műszaki és Gazdaságtudományi Egyetem, 2008.
14. Vicsi Klára, Szaszák György: Folyamatos beszéd szó- és frázisszintű automatikus szegmentálása szupraszegmentális jegyek alapján, II. rész: Statisztikai eljárás, finn-magyar nyelvű összehasonlító vizsgálat, III. Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Magyarország, 2005. pp. 360-370.