

NYELVELEMZÉS A SZÁMÍTÓGÉPES BESZÉDELŐÁLLÍTÁSBAN

Koutny Ilona

ELTE Általános és Alkalmazott Nyelvészeti Tanszék

A számítógépes beszédelőállítás vagy beszéd-szintézis feladata, hogy írott szöveget hangzó beszéddé alakítson át, azaz nyelvi jelekből akusztikus jeleket hozzon létre. A grafémák fonémákká történő leképezéséhez, mely a tényleges beszéd-szintézis kiinduló pontja, nem mindig elegendők fonetikai-fonológiai szabályok, mert bizonyos esetekben magasabb nyelvi szintek (morfológiai, szintaktikai vagy szemantikai szerkezet) határozzák meg a helyes kiejtést, a megfelelő prozódia kialakításához pedig a mondatok szerkezetének az ismeretén túl sokszor szemantikai és pragmatikai ismeretek is szükségesek. Ez azt jelenti, hogy jó minőségű beszéd-szintézis nem képzelhető el nyelvelemzés nélkül.

A jelen írásban körvonalazzuk, hogy hol a helye a nyelvelemzésnek a beszédelőállításban és milyen feladatokat kell ellátnia. A szótár nélküli szintézisre összpontosítva, mely komoly korlátozó tényező a nyelvelemzés számára, bizonyos megoldásokat veszünk sorra. Egy többnyelvű, szótár nélküli beszédelőállító rendszert tartva szem előtt, mint a MULTIVOX (Olaszy - Gordos - Németh 1992), példáinkat több nyelvből vesszük, de elsősorban a magyart (M), a németet (N) és az eszperantót (E) vizsgáljuk (további rövidítések: angol (A), francia (F), olasz (O)).

1. Gépi beszédelőállítás

A számítógépes beszédelőállítás sztenderd írott szövegből indul ki, melyben nemcsak betűk és írásjelek fordulhatnak elő, hanem számok és különböző egyéb karakterek (pl. % +), ezeknek a kimondásáról is gondoskodnia kell a rendszernek, erre szolgál az előfeldolgozási fázis (ld 3.).

A **betű--hang átalakítás** (melyet sokszor graféma--fonéma átalakításnak is hívnak) képezi a beszéd-szintézis-rendszer fő részét, a tényleges beszéd-szintézis az így kapott fonetikai címkékhez rendeli az adott nyelvre jellemző és a

beszédmegvalósítás eszközétől függő konkrét paramétereket. Megjegyezzük, hogy ez a fonémafogalom nem esik egybe a fonológusok által használtéval, a fonéma absztrakt fogalom, szintetizálni csak beszédhangot lehet. A kiejtést leíró szabályok lehetnek:

* környezetfüggetlenek, mint:

M: é --> [e], s --> [ʃ]; N: e --> [e], sch --> [ʃ], E: e --> [e], s --> [ʃ]

* környezetfüggőek, mint:

N: VsV --> [z], VsC --> [s], azaz intervokalikuss helyzetben az *s* ejtése [z], ha viszont mássalhangzó követi, akkor [s];

F: g{e/i} --> [ʒ], c{e/i} --> [s]

vagy akár az *n* kötelező lágyítása a lengyelben *i* előtt [n].

A kiejtés leírásához nem mindig elegendőek a fonetikai-fonológiai szabályok, hisz a magasabb nyelvi szintek felülírhatják ezeket a szabályokat, így például a morfológiai szerkezet más kiejtést rendelhet egy betűsorozathoz:

N: a *Häuschen* szóban az *sch* kiejtése [ʃç], mert az *s* után morfémahatár van; a *ge* kiejtése más [gə], ha előképző, mintha a szótóhoz tartozik: *geben* [gɛ] vagy *gelb* [gɛ].

A magyarban csak elvétve találunk ilyen példát (a *malacság* szóban külön ejtendő a *c* és az *s*), míg a németben ez a jelenség gyakori (ld. 4.).

A helyesírás konzervatizmusa miatt rendszerint akadnak minden nyelvben a szabályos kiejtés alól kivételt képező szavak, melyeket kiejtésükkel együtt egy kivételszótárba kell elhelyezni, pl. a *színház* szót rövid *i*-vel ejtjük. Az angol *ough* betűsorozatnak legalább 6 különböző kiejtését lehet megkülönböztetni: *ought* [o:t], *through* [θru:], *dough* [dou], *enough* [inʌf], *trough* [trɒf], *drought* [draut].

Az angolban többször, de a franciában is néha előfordul,

hogy egy szónak a kiejtését **szintaktikai szerepe** határozza meg:

A: record ['reko:d] főnévként és [ri'ko:d] igeként;

F: président [prezida] főnévként és [prezid] igeként.

Csak mondatelemzés tudja eldönteni, hogy az adott kontextusban éppen főnévi vagy igei szerepet játszik a szó. Egy teljes mondatelemző (ún. parser) csatolása nehézkessé tenné a beszédszintézis-rendszert, másrészt alig végezhető el valós időben, és nincs is mindig szükség rá. A mondat alapvető összetevőinek, a szintagmatikus szerkezetnek a megállapítása már sokat segít a helyes prozódia kialakításában (ld. 5.).

Az előbbieken említett szabályok és eljárások alkalmazása fordított sorrendben történik, azaz először a komplexebb, speciálisabb szabályok érvényessége kerül megvizsgálásra, és végül, ha semmilyen speciális szabály sem vonatkozik a kimondandó karaktersorozatra, akkor a legegyszerűbb környezetfüggetlen átíró szabályok lépnek működésbe. Most lássuk egy beszédelőállító rendszer építőelemeit (1. ábra)! A szótárt szaggatott vonallal kapcsoltuk a rendszerhez, jelezvén, hogy ennek nem mindig része. Az angol beszédszintézis elképzelhetetlen kiejtési szótár nélkül. A szótár a kiejtésen kívül rendszerint morfológiai és szintaktikai információkat is tartalmaz.



1. ábra A beszédszintézis folyamata

A természetes emberi beszédet az teszi igazán természetessé, dinamikussá, hogy értelmi és érzelmi hangsúlyok, illetve szünetek tarkítják a mondat típusa által meghatározott szupraszegmentális szerkezetet. Ezt a folyamatot szemantikai és pragmatikai ismeretek, valamint a helyzet irányítják. Ilyenfajta ismeretekkel nagyon nehéz ellátni a számítógépet, csak mesterséges intelligenciát alkalmazó (pl. kérdés-felelet) rendszereknél találhatunk próbálkozásokat ebbe az irányba, és ott is csak kiválasztott területekre korlátozva. A beszédelőállító rendszerek, hacsak nem egy mesterséges intelligenciával ellátott rendszer kimenetéhez kapcsolódnak, nem rendelkeznek ilyen irányú tudással. Az érzelmi hozzáállás, valamint pillanatnyi helyzet diktálta hangsúlyok megvalósítására egy számítógépes rendszer nem képes, mert ezek nem algoritmizálhatók, előre nem megadhatók.

2. Előfeldolgozás

Az előfeldolgozás feladata, hogy a szöveg karaktereit közvetlenül kimondható jelekké alakítsa. A billentyűzeten található speciális jelek feloldása nem okoz különösebb problémát, míg a rövidítéseké és mozaikszavaké igen, hisz elég változatos a skálájuk és egységes szabály nem alkalmazható rájuk. Illusztrációként a magyarból veszünk egy pár példát:

ált.: általános, db: darab, i. e.: időszámításunk előtt(i), sz.: számú, század

Problémát jelent a rövidítés végén található pont, mely néha egyúttal mondatvégi írásjel is lehet, továbbá a különböző értelmezés lehetősége. A mozaikszavak egy része szabályosan kimondható, míg más részüket betűzni kell: gyes, ÁFA, ENSZ, de IBM [ibeem], MTA [emtea], gmk [geemka]

A leggyakoribb rövidítéseket és mozaikszavakat listára kell felvenni a feloldásukkal vagy kiejtésükkel együtt, melyet

bővíteni lehet a várható alkalmazásnak megfelelően.

A számok kimondására algoritmus szolgál. Problémát megint csak a pont jelenti a magyarban és a németben, ahol ez a sorszámnevet jelöli, és néha egybeeshet a mondatvégi ponttal.

Míg a magyarban a sorszámnév kiejtése szabályosan generálható, a németben a generált alak végződését a hozzákapcsolódó főnév neme és szintaktikai szerepe határozza meg, valamint a névelő milyensége:

das 2. Haus [tsvajtə], de *im 2. Haus* [tsvajtən]

Így a megfelelő alak csak a mondatelemzés után érhető el. További problémát jelent a dátumok kiejtése, ahol a hónapok nevét kell behelyettesíteni a megfelelő sorszámnév helyére vagy a toldalékos alakok (a *20-án* kiejtésekor a *20-* helyébe a *huszadik* sorszámnév generálandó). A németben az év kiejtése is eltér a szokásostól: 1994: *neunzehnhundertvierundneunzig* és nem *tausendneunhundertvierundneunzig*.

3. Morfológiai elemzés

A morfológiai felbontás segítséget nyújt képzett szavak kiejtéséhez és a szintaktikai elemzés alapját képezheti. Az ismert MITalk (Allen et al. 1987) angol nyelvű beszéd-szintézis-rendszer is tartalmaz morfológiai elemzőt, így elegendő a szótöveket és a képzőket tárolnia (kevesebb, mint 10000 elemet), ezek kiejtését, továbbá bizonyos morfofonetikai szabályokat, mint az *s* rag zöngéssé válása zöngés végződés után: *dog + s* [dogz].

Az új szavak létrehozásában is sokszor a már létező morfémák vesznek részt. Morfémákra bontás után elkerülhető pl. a *hothouse* hibás ejtése, azaz a *th* egy hangnak tekintése.

A németben is bizonyos képzők -- az alapvető kiejtési szabályoktól eltérő -- meghatározott kiejtéssel rendelkeznek, pl. *be-* [bə], *-tion* [tsjon], illetve hangsúlyuk is kötött: vannak hangsúlyos (*an, auf, aus, ...*) és hangsúlytalan (*be, ge ... ver*)

előképzők (további példákat ld. Wothke 1991, a német morfológiai elemzéshez pedig Pounder - Kommenda 1986).

4. Szintaktikai szerkezet és szupraszegmentális szerkezet

Két megközelítés lehetséges. Az egyik szerint a szintaktikai szerkezet az alapja a szupraszegmentális szerkezetnek. A másik szerint a két szerkezet között nincs egyértelmű megfelelés, hisz a szintaktikai szerkezethez képest az intonáció kevésbé változatos, de ennek ellenére előfordul, hogy ugyanahhoz a szintaktikai szerkezethez többfajta intonáció rendelhető, azaz a szemantikai tartalom nem mindig a szintaktikai komponensen keresztül jut el a prozódiaihoz, hanem közvetlenül is befolyásolhatja azt (Monaghan 1993).

Gondoljunk a *Péter levelet ír.* típusú mondatokra, ahol egy semleges és egy fókuszos - azaz a *levelet* hangsúlyozó - ejtés is lehetséges. Csak egy szövegszintű elemzés próbálhatná meg eldönteni, hogy az adott helyzetben éppen melyikről van szó. A magyarban az eldöntendő kérdést is a kijelentéstől csak az intonáció különbözteti meg, de itt a mondatvégi írásjel egyértelműsít.

Ha egyértelmű megfelelés nem is, de bizonyos összefüggés azért fennáll a mondat szerkezete és szupraszegmentális struktúrája között. A szintagmákon belül szorosabb a kapcsolat, mint a szintagmák között, ez nyomon követhető a hangsúlyokban, dallamban, a lehetséges szünetekben, bizonyos időviszonyokban (pl. nyújtás a ritmikai egység végén) és más jelenségekben, mint például a franciában megnyilvánuló *liaisonban* (*hangkötés*: a *de bons enfants* főnévi csoportban az amúgy néma többes számú melléknévi végződés z alakban összeköti a melléknévet a rákövetkező főnévvel).

A francia vagy az angol ritmikai egység vagy fonetikai szó, mely közelebb áll a szintagmához, egy fő hangsúllyal rendelkezik (a franciában ez az utolsó szótagra esik). A ritmikai

egységen belül folyamatos az ejtés, de közöttük lehet szünet.

A következő pontban a mondatok szintagmaszintű szerkezetének a feltárásával foglalkozunk, mely szerintünk is bizonyos korrelációt mutat a semleges szupraszegmentális szerkezettel. Nem áll módunkban kitérni az intonáció és prozódia konkrét megvalósulására különböző szintaktikai konstellációkban, erre lásd Varga 1989, valamint É. Kiss 1988 cikkeit a magyart illetően.

Mint már említettük, gépi beszédelőállításról lévén szó, le kell mondanunk az intonáció egyik legfontosabb funkciójáról, az érzelmkifejezéséről. Tudjuk, hogy az intonáció bizonyos esetekben az ellenkezőjére fordíthatja a verbális nyilatkozat értelmét és sokat számít a beszéd természetes megítélésében. A rossz prozódia zavarhatja az érthetőséget is. Az intonáció részt vesz a kommunikatív jelentés kialakításában, a téma-réma szerkezetet az intonáció alátámasztja, illetve általa jut kifejezésre (l. Péter 1991). A továbbiakban a semleges alapproforma kialakítását kívánjuk mondatelemzéssel segíteni.

5. Mondatelemzés

A szintaktikai elemzés (parsing) során a mondat lineáris sorából előállítjuk a hierarchikus szerkezetét, összetevői vagy az elemei között fennálló függőségi viszonyok alapján, amelyet fa-alakban lehet ábrázolni. A teljes elemzés a szótári elemek morfológiai, szintaktikai és szemantikai információira alapul rendszerint. Például az MITalk rendszer a morfológiai elemzés kimenetére és a morfémák szótári információira (mint szófaj) támaszkodik a szintaktikai elemzéskor, de nem törekszik kimerítő elemzésre. Megelégszik a mondat főnévi és igei csoportokra való bontásával, így elkerül bizonyos kétértelműségeket (Allen et al. 1987).

A következőzetes (!) központosítás is segít nagyobb egységek különválasztásában. A magyarában a vesszőt, pontosvesszőt

vagy kettőspontot feltétlenül határformának kell tekinteni, mely szünetet von maga után és befolyásolja az intonációs kontúrt.

Általában a mondat strukturálisan is lényeges elemei hordozzák a hangsúlyokat. Ezek automatikus kijelölése szótár nélkül igen nehéz, de a hangsúlytalanok legalább egy részének a felderítése gyümölcsözőbbnek ígérkezik. A funkciószavak (kötőszavak, elöljárók, determinánsok stb.) nagy része hangsúlytalan. Egy kezdetleges mondatelemzés, mely egy kis szótár alapján azonosítani tudja a funkciószavakat, sokat javíthat a gépi beszéd természetességén (O'Shaughnessy 1989). O'Shaughnessy megpróbálja továbbá a szintagmahatárokat is detektálni az angolban a funkciószavak és szórendi megkötések segítségével.

A szótár nélküli megközelítés jól alkalmazható az eszperantóra, ezért választottuk modellként.

Modell: eszperantó mondatelemzés

Az eszperantó mondatok elemzése elvégezhető szótár nélkül, mivel elemei egyértelműen leírhatók formális jegyek alapján, pl. főnév végződése *o, oj, on, ojn* (az utóbbi két alak már tárgyestű), a melléknév *a, aj, an, ajn*. Egy főnévi csoport (NP) szerkezete a következő:

NP: (DET) (NUM) (AP*) N (AP*) vagy NP: NP PP

PP: PRP NP

AP: (ADV) ADJ, AP: AP PP, AP: PP AP

() jelöli az elem fakultatív voltát, míg a * az ismételhetséget. A PRP (elöljáró), DET (determináns) kategóriák felsorolhatók, NUM (számnév) pedig szabállyal megadható. Az egyszerű melléknévi csoport, mely csak egy melléknévből áll, belesimul a főnévi csoportba ejtés szempontjából, de az összetettebbek önálló egységet alkotnak.

Igei csoporton (VP) itt az ige és az ejtés szempontjából feltétlenül hozzá simuló elemeket értjük.

VP: (PRN) V (Vi) (PRN) VP: (PRN) V (Vi) Adj/N

A PRN (névmás) kategória zárt osztály, az ige (V) végződése *as, is, os, us, u*, a főnévi igenévé (Vi) pedig *i*. A VP-ben megjelenő Adj-on az állítmánykiegészítő vagy pedig az összetett állítmány igenévi része értendő. Így egy balról--jobbra haladó "*bottom-up*" parser könnyen elő tudja állítani például a következő mondat szerkezetét (melynek jelentése: Négy nagyon szorgalmas diák el akarja ezt olvasni az Akadémia könyvtárában):

Kvar tre diligentaj gestudentoj volas legi tion en la biblioteko de la Akademio.
 NUM ADV ADJ N V Vi PRN PRP DET N PRP DET N
 NUM AP N VP PRP NP PRP NP
 NP VP PP PP

Ez a felbontás már megadja a prozódia szempontjából lényeges határokat. A csoportokon belül kötés és a funkciószavak hangsúlytalanítása mehet végbe (Koutny 1990).

A PP-k kötése más elemekhez szemantikai értelmezés nélkül hibás szerkezeteket hozhat létre, pl. a *mi pruntas libron de amiko* esetében, ha kötjük a *de amiko*-t, az előtte levő NP-hez, akkor azt jelenti, hogy *kölcsönveszem a barátom könyvét*, míg enélkül: *kölcsönveszem a könyvet a barátomtól*.

Magyar mondatok szintagmatikus szerkezete

Itt csak az egyszerű magyar mondat szerkezetét kíséreljük meg nagy vonalakban elemezni. É. Kiss szerint a mondat egy logikai alany szerepű összetevőre és egy logikai állítmány szerepű igei csoportra tagolódik (Kiefer 1992). Az előbbi a topik (téma), az utóbbi a komment (réma), melynek első fő összetevője kötelezően hangsúlyos. Ezt a helyet a fókuszba helyezett elem, vagy annak hiányában az ige tölti be. A fókusz és az ige egy fonetikai szóba tartozik. Még bizonyos kvantorok állhatnak a fókusz előtt. Lássunk most egy egyszerű mondatot!

Jancsi a barátait többször is vacsorára hívta az otthonába, családtagjaikkal együtt.

topik (T) kvantor (Q) fókusz (F) ige (V) többi elem (N)

A T, F és N rendszerint egy vagy több főnévi csoportból, esetleg határozószókból áll, a Q pedig felsorolható elemekből, így a feladat durván az NP-k meghatározása. Az NP-ken belül szünet nélküli az ejtés, még az NP-k között lehet szünet. Feltételezzük, hogy egy egyszerű NP általános szerkezete a következő:

NP: (DEM) (ART) (PRN/N) (QUA) (NUM) (AP*) N
(POSTP)

pl. ez az én bármelyik 3 kedves könyvem
mind a 3 kedves könyvem
minden 5. nagy ház után

AP: (ADV) ADJ

DEM: ez/az és ragozott alakjaik, mind

ART: a(z), egy

PRN: én, te,..., valaki, ...

QUA: minden, mindegyik, valamelyik, ...

NUM: tőszámnevek, sorszámnevek, sok,....

POSTP: alatt, át, előtt, ...

Így a barátom egyik könyve egy NP-be tartozik, de a barátom egy könyve egyelőre két NP-be, ugyanígy a bővítményes melléknévi igenevet tartalmazó NP-k, mint az előadáson részt vevő hallgatók (hisz ilyen is előfordulhat hasonló szerkezetben: az előadáson érdeklődő hallgatók ültek). Ezek egyesítése egy NP-be egy további lépés.

Az *is* ragadhat még az NP végére hangsúlytalanul. Az *és*, *vagy* pedig elválaszthat két NP-t.

Formai jegyek alapján nem lehet a szavak szófajára következtetni, mint az eszperantóban, így a funkciószavakat (DEM, ART, PRN, QUA, NUM, POSTP) érdemes felderíteni,

hisz ezek egyszerűen felsorolhatók. A főnévi szintagmát a ragozott főnév vagy a névutó zárja. A rag fontos szerepet tölt be a szintaktikai funkciók jelölésében, azaz segítségével állapítható meg a mondat elemeinek egymáshoz való viszonya. Így a teljes mondatszerkezet feltárása csak a szavak szófaját tartalmazó szótárhoz való fordulás és morfológiai elemzés után lehetséges.

Az ige központi szerepet tölt be a mondatban, a fő bővítmények (alany, tárgy, különböző vonzatok) hozzá kapcsolódnak. Kiejtés szempontjából most csak egy szűkebb - feltétlenül egy fonetikai szót alkotó - VP-t tekintünk, még pedig a komplex igét (igekötő (PRF) + ige, segédige (AUX) + ige. V + Vni, V<past> + volna), valamint az ige előtt álló fókusz vagy módosító elemek és az ige csoportját, pl:

szeretnék olvasni, olvastam volna, el fogom olvasni,

el kell olvasnom, nem olvasom el, a könyvet olvasom el.

Ezen szerkezetek felismeréséhez szükséges az igék azonosítása ragjaikkal együtt, de a segédigéket, az igekötőket és módosítókat fel lehet sorolni. Amennyiben az igekötő az ige után található, az ige előtti csoport fókuszának minősül, s ennek megfelelően kap hangsúlyt, ha pedig előtte, akkor maga válik hangsúlyossá.

Egy jobbról-balra haladó elemző tudná a magyar mondat szerkezetét felderíteni a ragokat használva fogódzkodóként, és megengedve a visszalépést a helyes struktúra megtalálása érdekében.

Befejezésként Vargát (1989) idézem: "*A felszíni szintaxis a magyarban a legtöbb esetben meghatározza azokat a helyeket, ahol valamilyen intonációs határformát alkalmazni kell.*" *A performancia szabadsága az, hogy a beszélő választhat ott, ahol több határforma lehetséges.* A beszéd-szintézisnél a választást előre meg lehet határozni, vagy pedig véletlenszerűen a gépre hagyni a döntést.

Az eddigi rövid bemutatás csak vázolta a problémákat, jelezve, hogy milyen szintű nyelvelemzés tudná azok megoldását elősegíteni. Ezek alkalmazása a konkrét beszéd-szintézis rendszer lehetőségeitől és az adott nyelv fonéma-graféma megfelelésétől függ.

Irodalom

Allen, J. - Hunnicutt, M.S. - Klatt, D.: From text to speech. The MITalk system. Cambridge University Press, Cambridge, 1987.

Bailly, G. - Benoit, C. (eds): Talking Machines. Theories, Models, and Designs. North-Holland, Amsterdam 1992.

É. Kiss Katalin: Még egyszer a magyar mondat intonációjáról és hangsúlyozásáról. NyK LXXXIX, 1988, 1-52.

Kiefer Ferenc (szerk.): Strukturális magyar nyelvtan. 1. kötet: Mondattan. Akadémia Kiadó, Budapest, 1992.

Koutny Ilona: Gépi beszéd-előállítás és alkalmazása a magyar és az eszperantó számítógépes oktatásában. 1990. Bölcsészdoktori disszertáció, ELTE

Monaghan, A.I.C.: Parsing unrestricted text: a multiphase approach. EUROSPEECH '93. Berlin, 1993.

Olaszy G. - Gordos G. - Németh G.: The MULTIVOX multilingual text-to-speech converter. Talking Machines. 1992, 385-411.

Péter Mihály: A nyelvi érzelmekifejezés eszközei és módjai. Tankönyvkiadó, Budapest, 1991.

O'Shaughnessy, D.: Parsing with a small dictionary for applications such as text-to-speech. Computational Linguistics 15/2, 1989, 97-108.

Pounder, A. - Kommenda, M.: Morphological analysis for German text-to-speech synthesis. Proc. of 11th COLING. Bonn, 1986.

Varga László: Szintaktikai szerkezet és intonációs tagolás a magyar mondategészben. ÁNyT XVII, 1989, 241-278.

Wothke, K.: Letter-to-Phone Rules for German. IBM, Heidelberg, 1991.