

RÉSZLETES KUTATÁSI TERV

1. A magyar nyelv szempontjából központi fontosságú megértési tényezők kísérleti elemzése

1.1 Alaktani feldolgozás és megértés. Az agglutináló alaktan szerepe és különlegességei a megértési folyamatban

A magyar nyelv agglutinatív szerveződésének megfelelően kitüntetett szerepet játszhat az alaktani feldolgozás elméleteiben. Három mozzanatot tervezünk vizsgálni, belső fejlesztésű már meglévő szoftvereinket tovább alakítva. A mai pszicholingvisztikában nagy viták folynak a folyamatos megértés közbeni feldolgozási döntések elemző vagy egészes jellegeről. Vajon az (1) szó megértése a (2) vagy a (3) lépéssornak megfelelően megy-e végbe?

(1) *barátaimat*

(2) ANALITIKUS barátaimat → BARÁT-A-I-M-AT

(3) EGÉSZLEGES barátaimat → BARÁTAIMAT

A feldolgozásról eddig is sok adatunk van (lásd a Pléh és Lukács, 2001 szerkesztette kötetben), amelyek hibrid modellek mellett szólnak: rendelkezésünkre áll egy analitikus eljárás, ezt azonban modulálja a gyakoriság, a nem áttetsző jelentés, és az alaktani kivételesség, mindhárom az egészes hozzáférést preferálva. Tervezett új kutatásaink három szempontból jelentenek továbblépést.

Összevetjük az akusztikus és az olvasott szövegfeldolgozást

Az eddig használt lexikai döntés (A (4)-ben szó-e pl. *a, b, c, d*)

(4) *a. barátaimhoz*

b. baréaimhoz

c. barátaimnek

d. barátaimguz

helyzetek mellett modalitásközi és modalitáson belüli (akusztikus-akusztikus és vizuális-vizuális) előfeszítési elrendezésekben vizsgáljuk azt, mennyire analitikus és mennyire egészes a feldolgozás az említett ingersajátosságok függvényében.

A helyzet logikája és a várt eredmények a következők:

(5)	ELŐFESZÍTŐ SZÓ	CÉLSZÓ	KIOLVASÁSI IDŐ
	<i>házasság</i>	<i>kert</i>	lassú
	<i>házakban</i>	<i>kert</i>	gyors

Alapvető kérdésünk az, hogy az analitikus-holisztikus feldolgozás mennyire függ a gyakoriságtól, a töminta produktivitásától, a szemantikai áttetszőségtől és a bemutatási módtól, mint a Marslen-Wilson és mtsai. (1994) elindította programban is előtérbe került. Összevetjük izolált és mondatban álló komplex szavak feldolgozását, elsősorban torzított (például kattánójajjal maszkol) akusztikus anyagok segítségével. Mindezek a kísérleti helyzetek kidolgozásra kerülnek különböző életkorú gyermekcsoportokra is. Ez annak tisztázása érdekében releváns, hogy milyen szerepe lehet az olvasás elsajátításának az analitikus megértési eljárások uralkodó válásában a magyarban.

1.2 A vonzathozzárendelés algoritmusai, morfológiai és szintaktikai meghatározói, a kétértelműség fajtái és megjelenése a megértés során.

A magyar nyelvben különösen áttetsző és izgalmas kérdés, ahogyan a mondatmegértés során kölcsönhatásba lépnek a mondattani szereprendszerek és a főnévi csoportokon az alaktani jelölők. Kísérleteink során ezeket az interakciós viszonyokat szeretnénk tisztázni, beleértve a szórend, a fogalmi viszonyok és a vonzathozzárendelés kapcsolatát.

Szavankénti olvasási, torzított hanganyagot használó és váratlan felidézési instrukciós helyzeteket kialakítva vizsgáljuk, hogy milyen szerepe van az igékből keletkező elvárásoknak és a mondattani szerepre nézve a ragok által hordozott információnak a megértésben. Milyen idői menetben történik az alaktani és a vonzatokra vonatkozó mondattani információ összekapcsolása. (6) és (7) összehasonlítása mutatja az egyik alaphelyzetet

- (6) *A rendőr a megoldáson töprengett.*
(7) *A rendőr töprengett a megoldáson.*

Vajon a vonzat (megoldáson) segíti az ige, vagy fordítva, az ige segíti a vonzat feldolgozását? (8) azt is bemutatja, hogy itt jellegzetes hatása lehet szemantikai illetve kategorikus megkötéseknek. (8c) kétértelmű, már ami a ragozott főnév szerepét illeti.

- (8) a. *Anna a széken ült.*
(8) b. *Anna a kérdésen gondolkodott.*
(8) c. *Anna a hajón gondolkodott.*

Gervain Judit olvasási kísérletei után most kifinomultabb eljárásokat használva a pontosabb idői viszonyok feltárására törekszünk.

Ennek során olyan kísérleteket is tervezünk, ahol szembeállítjuk egymással a különböző lehetséges komponenseket, sértési helyzetekben, mint (9) mutatja, s a reakcióidők mellett (el kell dönten, jó-e a mondat) a különböző latenciájú agyi elektromos válaszok elemzésével következtetünk a megértési lépések sorrendezésére.

- | | |
|--|------------------|
| (9) a. <i>Az utcaseprő lekotorta a havat.</i> | KORREKT |
| (9) b. <i>Az utcaseprő lekotornak a havat.</i> | MONDATTANI HIBA |
| (9) c. <i>Az utcaseprő lekotorta a hót.</i> | ALAKTANI HIBA |
| (9) d. <i>Az utcaseprő lefölözte a havat.</i> | PRAGMATIKAI HIBA |

A nemzetközi irodalomban a különböző sértéseknek megfelelő kiváltott agyi válaszkomponensekről, s ezek idegrendszeri topográfiájáról számos elmélet s vita létezik, a 150 és 600 ms közötti különböző idői ablakokban. A magyar kísérletek különleges érdekessége az lehet itt, hogy módot adnak több komponens világosabb szembeállítására, éppen a gazdag alaktan révén.

1.3 A mondatközi integráció és a szövegértés pontos idői rendszere.

A klasszikus vizsgálatok azt mutatják, hogy a mondatok formai jegyeire csak egészen rövid ideig emlékszünk, s csak a mondatok tartalmát tartjuk meg a hosszú távú emlékezetben. Az eljárás során ingermondatokat adunk a személyeknek (10a), majd leteszteljük, mennyire pontosan emlékeznek arra, hogy hogyan hangzott a mondat.

- (10) a. *Küldött erről egy levelet Galileinek, a nagy olasz tudósnak.*

Az (a) mondat meghallgatását követően a személyeknek meghatározott késleltetési idő elteltével azt kell eldönteniük, hogy az alábbi mondatokat hallották-e.

- (10) b. *Egy levelet küldött erről Galileinek, a nagy olasz tudósnak.* (szórend változás)
(10) c. *Küldött erről Galileinek, a nagy olasz tudós, egy levelet.* (tagmondatrend változás)
(10) d. *Galilei, a nagy olasz tudós, küldött neki erről egy levelet.* (jelentés változás)

A klasszikus eredmény szerint (b) és (c) mondatot már 40 ezredmásodperc után sem tudjuk megkülönböztetni (a) mondattól, a jelentésében megváltoztatott (d) mondatot viszont el tudjuk különíteni. Mindez arra utal, hogy a mondatok formai jegyeit csak egészen rövid ideig tároljuk, csupán a mondatok jelentése az, ami a hosszú távú emlékezetben megmarad. A sajátosan érdekes mozzanat az, hogy a magyarban a szórend mint tényező lehet jelentés irreleváns (c) illetve jelentés releváns a fókuszba emelés révén (b). Kérdés, hogy a fókuszba emelés, mint jelentést hordozó formai változás, megőrződik-e az emlékezetben.

1.4 A munkaemlékezet működő rendszere egy agglutinatív nyelvben.

A munkaemlékezet alrendszerének kulcsszerepe van a nyelvi folyamatokban (Baddeley, 2003). A mondat és szöveg megértés valamint a társalgási folyamatok elképzelhetetlenek a rövid idejű tár(ak) nélkül. A nyelvhasználat, nyelvi megértés egyéni különbségeinek hátterében számos esetben a munkaemlékezet különböző komponenseinek korlátozott kapacitása áll (Caplan és Waters, 1999). Ez nemcsak a normál nyelvi működésre, hanem a nyelvpatólogiákra, és a nyelvfjlődésre is igaz.

Az alaktani folyamatok és a munkaemlékezet kapcsolata eddig feltáratlan terület. A kutatás célja, hogy pontosan feltérképezze a munkaemlékezet komponenseinek szerepét a különböző alaktani folyamatokban, valamint új vizsgálati eljárásokat dolgozzon ki. Vizsgálatainkban célunk a verbális munkaemlékezet klasszikus

mérőeljárásainak átalakítása olyan formába, hogy az magyar alaktani rendszer működéséről legyen képes adatokat szolgáltatni. A szóterjedelem feladat szuffixumos változatával folytatott előkísérleteink szerint a szuffixumos alakok munkaemlékezeti tárolása lényegesen nehezebb a tövek tárolásánál. Minél több szuffixum van a szólista szavain annál több munkaemlékezeti kapacitás szükséges.

Az elővizsgálatok szerint tehát az eljárás segíthet a fenti kérdések megválaszolásában. Vizsgálatainkban a következő célokat tűztük ki:

1. Több új munkaemlékezeti mérőeljárás kidolgozás, mely pontos adatokkal szolgál a magyar alaktani működésről.
2. A magyar alaktani működés (szegmentálás, elemzés, jelentésintegráció) folyamatainak részletes leírása.
3. A szabályos-rendhagyó, képzett-ragozott, valamint a különböző típusú szuffixumok eltérő munkaemlékezeti erőforrás igényének feltérképezése.
4. A munkaemlékezet komponensek szerepének feltárása az alaktani elemzésben.

2. Elsajátítás

2.1. Magyar anyagra speciális eljárásokat dolgozunk ki a spontán szövegeken alapuló fejlődésemelés számítógépes segítésére a CHILDES adatfeldolgozó rendszerben

Az amerikai eredetű CHILDES adatfeldolgozó rendszer honosítása és terjesztése a kognitív tudományokkal foglalkozó kutatók részére alapvető feladat. A CHILDES (Child Language Data Exchange System. MacWhinney, 2003) egy elektronikus többnyelvű gyermeknyelvi adatbázisból és az adatok számítógépes feldolgozását elősegítő programcsomagból áll. A rendszer különlegessége, hogy spontán nyelvi anyagra (korpuszra) épül, amelynek elérhetővé tétele új dimenziót nyit meg a kognitív kutatásokban. A kísérleti kutatási módszereket kiegészítve, a spontán nyelvi anyag vizsgálata például lehetővé teszi új problématerületek feltárását, a kognitív fejlődési folyamat nyomon követését és a természetes környezet hatásának vizsgálatát, beleértve a szülői/gondviselői nyelvhasználat megfigyelését. A korpusznyelvészet azonban mind gyűjtési, mind feldolgozási módjaiban jelentősen eltér a kísérleti alapú kutatási módszerektől. A természetes nyelvi korpusz jelenleg Magyarországon a nyelvtechnológia területén elterjedt eszköz, például a szótárkészítésben és a gépi fordításban. A CHILDES vívmánya egy olyan rendszer megteremtése, ahol a nyelvtechnológia módszerei ötvöződnek a kísérleti pszicholingvisztika céljaival.

A rendszer honosítása jelenleg folyamatban van kutatócsoportunkban: a magyar nyelvi felhasználáshoz szükséges alapfunkciók megvalósításra kerülnek a jelen pályázat megkezdése előtt. A továbbiakban célunk a rendszer ismertetése a hazai szakmai közönséggel és mind a korpusz, mind az adatfeldolgozást segítő funkciók (algoritmusok) folyamatos bővítése a kognitív tudomány felmerülő kérdéseit szem előtt tartva. Az eddigi kutatómunkából rendelkezésre álló magyar gyermeknyelvi korpuszt a CHILDES leírásintaxisának megfelelően konvertáljuk. Terveink szerint a Kognitív Tudományi Tanszék jövőben felvett adatait ezekhez a specifikációkhoz igazodva fogjuk átírni.

A korpusz elemzéséhez és statisztikai információ kinyeréséhez elengedhetetlen a szöveg kódolása, vagyis a szavak/szócsoportok releváns címkézése. A CHILDES programcsomagja lehetőséget nyújt a címkézés automatizálására, ehhez azonban a célnyelv nyelvtani szerkezeit leíró szabályrendszerre van szükség. A magyar nyelvben lényeges információ egyrészt a szavak szófaji meghatározása és morfológiai elemzése, másrészt egy-egy mondaton belül a különböző szintaktikai szerepű szócsoportok lineáris elhelyezkedése a mondat állítmányához viszonyítva. Elsődleges tervünk tehát egy morfológiai elemző telepítése a CHILDES programnyelvén és egy szócsoport felismerő és szintaktikai szerep meghatározó komponens kidolgozása. Az így nyert annotált korpusz lehetőséget ad a magyar nyelvsajátítás tipikus menetének átfogó leírására, valamint a pszicholingvisztika számos kulcstémájának kísérleti anyagot kiegészítő vizsgálatára.

A gyermeknyelv, és ezen keresztül az emberi megismerés, szemantikai motivációinak korpuszalapú vizsgálatához azonban további fejlesztésre lesz szükség. Ehhez meg kell jelölnünk az egyes szavak, szócsoportok jelentéstípusát. A fél-automatikus szemantikai címkézés a morfológiai és szerkezeti annotációra építve speciális szótárak létrehozásával valósítható meg. A névszói szócsoportok mondatszintű szemantikai szerepe például kiolvasható egy vonzatkerettárból, melyben egy-egy ígéhez (régenshez) tartozó vonzatok morfológiai (és bizonyos esetekben elhelyezkedési) specifikációit szemantikai szerepükkel párosítjuk. A szemantikai kódolás elősegíti a kutatás céljaihoz releváns szövegrészek kigyűjtését és rendszerezését. Korai gyermeknyelvi anyagban azonban a fent vázolt gépi szemantikai címkézés csak irányadó szerepet kaphat, hiszen a nemzetközi kutatások arra mutatnak, hogy a gyermeknyelvben a szemantikai kategóriák formai megjelenítése nem feltétlenül tükrözi a felnőtt nyelv szabályait. Az automatikus annotációt tehát kézi kiegészítés kell hogy kövesse.

A CHILDES adatfeldolgozást elősegítő programcsomagja nyelvtől független és mindenki számára elérhető minden elterjedt operációs rendszerben. Használata alapszintű programozási készséget igényel. Célunk az iskolánkban tanuló diákok kiképzése a programok használatára, s specifikus hipotézisek vizsgálata, például az korai szórend grammatikai, szemantikai és pragmatikai alapjairól. Ide tartozik a morfológiai szótípusok elsajátítása, az elsődleges nyelvtani kategóriák szerveződése, a szórendi kötöttségek alakulásának vizsgálata is.

2.2 A korai nyelvi fejlődés kísérleti vizsgálata.

2.2.1 Nyelvfelődés és szociális kompetencia a kétnyelvűségben

Két nyelv korai és egyidejű elsajátításának vizsgálata értékes adatokat szolgáltat a nyelvfelődés folyamatáról, univerzális illetve nyelvspecifikus aspektusairól. Az itt vizsgálatra javasolt kérdéskörök - 1. a kétnyelvű lexikon és szintaxis fejlődése és szerveződése, 2. a tudatelmélet szerepe a kétnyelvű szociális és nyelvi interakcióban - közelebb vezetnek a nyelv kognitív architektúrájának megértéséhez.

Az első problémakör tanulmányozása során azt vizsgáljuk, hogyan különít el a kétnyelvű gyermek két mentális objektumot (két nyelvet), hogyan történik ezek szimultán elsajátítása, illetve hogyan változnak a két nyelv interakciójának mintázatai a fejlődés folyamán. A korai szótanulás fő elméleteit próbáljuk a kétnyelvű elsajátításra kiterjeszteni. Kérdés, hogy kielégítő magyarázatot szolgáltatnak-e ezek az elméletek – pl. a Kölcsönös Kizárásról illetve a tudatelméleten alapuló (Bloom, 2001) - a kétnyelvű szóelsajátításra? Hogyan alkalmazzák a kétnyelvű gyermekek a Kölcsönös Kizárás elvét, amely szerint minden tárgynak csak egy neve lehet az új szavak tanulása során miközben ők eleve két címkét kell megtanuljanak? Hogyan és milyen mértékben használják fel a kétnyelvűek a korai tudatelméleti képességeiket (mások intencióinak, megértését) a nyelvek megkülönböztetésében illetve a szótanulásban?

A második kérdés a nyelvfelődés és a mások intencióinak, hiedelmeinek megértése közötti kapcsolat tanulmányozására irányul: melyek azok a kognitív előfeltételek, amelyek megalapozzák a nyelvsajátítást, illetve milyen szerepet játszik a nyelvfelődés más kognitív folyamatok fejlődésében. Az újabb kutatások eredményei korrelációt mutatnak ki a nyelvfelődés és a mások mentális állapotainak megértése között. Az egyik kutatási irányzat szerint (Bloom, 2001, Gopnik és Meltzoff, 1997, Tomasello, 2001) a tudatelmélet, a mások mentális állapotainak, kommunikatív intencióinak bizonyos szintű megértése alapvető feltétele a szójelentések elsajátításnak. Amikor a szülő egy új szót tanít a gyermeknek, a gyermek csak úgy tanulhatja meg, ha megérti a szülő kommunikatív intencióját és a szó referensét. Más kutatások viszont arra találtak adatokat, hogy a nyelvtan bizonyos aspektusai (a feltételes mellékmondatok megértése: pl. *Béla azt mondta, hogy Éva alszik, de ő tulajdonképpen a tévét nézi.*) kritikusak a klasszikus értelemben használt tudatelmélet kialakulásában.

Kétnyelvű gyerekeket vizsgálva értékes adatokat nyerhetünk ebben a témakörben. Egy korábbi kutatásunkban (Kovács Ágnes) három éves egynyelvű és kétnyelvű óvodások tudatelméleti képességeit hasonlítottuk össze. Feltételezéseink szerint a két nyelv szimultán használata pozitív hatással lehet a tudatelmélet fejlődésére tekintve, hogy a kétnyelvűeknek a beszélgetőpartnerük nyelvi tudása függvényében kell módosítaniuk kommunikációs viselkedésüket. A megfelelő nyelv kiválasztása egy bizonyos szintű betekintést feltételez a mások elméjébe (milyen nyelveket ismer a másik személy). Ezek olyan megszorítások, melyek a felnőttekhez hasonló szintű tudatelméleti képességek korábbi kialakulásához vezetnek a kétnyelvű óvodásoknál. Az eddigi eredmények szerint a kétnyelvű három éves óvodások jelentősen nagyobb hányada mutat jó teljesítményt a tudatelmélet próbákon az egynyelvűekhez viszonyítva. A tervezett kutatások arra szeretnének választ találni, hogy mit jelent ez a kétnyelvű előny a szociális kompetenciák fejlődésére nézve, illetve, hogy milyen mechanizmusok támogatják (lásd a kontrollfolyamatok illetve a nyelvtani képességek egyre hangsúlyozottabb szerepét).

2.2.2 A szuffixumok absztraktságának és általánosíthatóságának vizsgálata

A nyelvsajátítást paraméterbeállításként kezelő és a fejlődés során az absztrakt nyelvi kategóriák korán megjelenő tudását feltételező generatív elméletekkel szemben az elmúlt 10 év gyereknyelvi kutatásainak és új nyelvészeti irányzatainak alapján egy gyökeresen új alternatíva jelent meg arra nézve hogy hogyan tanulja meg a gyerek az anyanyelvét. A felhalmozódott eredmények szintézise Michael Tomasello nevéhez fűződik (pl. Tomasello, 2003), aki kutatócsoportjában számos újabb vizsgálatban azt mutatta ki, hogy a korábbi feltételezésekkel szemben az absztrakt konstrukciók produktivitása kezdetben nagyon korlátozott: a gyerekek új főnevet viszonylag szabadon használnak új konstrukcióban, de ígét nem nagyon, intranszitiván megtanult új ígét transzitiván még akkor sem, ha ez lehetséges az ige jelentése alapján. Érdekes kérdés, hogy csak a szintaktikai konstrukciók produktivitása ilyen szűk kezdetben, vagy a lokálisabb és ezért talán könnyebb, ragok mint absztrakt konstrukciók vagy képzők produktivitása is ennyire kicsi eleinte, főleg olyan nyelvekben mint a magyar. A főnevek és az igék is jobban összehasonlíthatók így, mivel a főnevek nemcsak az igékkel, hanem a saját toldalékaikkal is konstrukciót alkothatnak. Ezt a kérdést vizsgálnánk új igék és ragozott alakjainak tanításával (*Nézd! A vag! vagy Nézd, mit csinál! Ezt glorpolásnak hívják! vagy Nézd mit csinál! Glorpol!*), majd

megpróbáljuk kiváltani egy másik olyan másik rag használatát, amelyiket az előzetes megfigyelés alapján a gyerek létező szavakkal már használ: pl. (pl. *Mi történik, mi történt* kérdések, amelyek nem tartalmazzák a kritikus ragot, egyébként: *Mit, mivel, mihez, kié,* stb: *Mik ezek? Mivel ütögetem az asztalt? Mitől ijedt meg a maci? Mit csinál? Mit csinálok? Mit csinálsz? Mit csinálnak? Tegnap mit csinált? Mit fogunk csinálni? Mond meg neki, hogy mikoljon).*

2.2.3 A prozódiai és statisztikai információ szerepe a szórend és az agglutinatív jelleg elsajátításában

Az Elvek és Paraméterek modell (pl. Chomsky 2000) szerint az anyanyelv elsajátítása a paramétereknek az adott nyelvre jellemző beállításainak meghatározására redukálódik. A szórendet meghatározó egyik alapvető paraméter például azt dönti el, hogy egy mondatrész és a vonzata milyen sorrendben jelennek meg a mondatban. Az angolban például az ige megelőzi a tárgyát (*eat an apple*), az előljárószó a főnevet (*under the table*). Ezzel szemben a japánban például, sőt a legtöbb esetben a magyarban is, a tárgy előzi meg az igét (*almát eszik*), a főnév a névutót (*az asztal alatt*). A csecsemőnek tehát az a feladata, hogy beállítsa ezt a paramétert a két lehetséges állás, a „vonzat elől” (magyar, japán, baszk, török stb.), illetve a „vonzat hátul” (angol, francia, olasz stb.) egyikébe.

Ahhoz azonban, hogy a gyermek az elvont szintaktikai paramétereket be tudja állítani, össze kell tudnia kötni azokat az általa hallott konkrét nyelvi adatokkal: e „kötési probléma” (Pinker 1984) feloldására dolgozták ki az úgynevezett „csizmahúzási” elméleteket. Az elméletek egy csoportja (Nespor et al. 1996, Morgan és Demuth 1996) a gyermeket érő nyelvi inger hangalakjában véli felfedezni a csizmahúzás jelét. Nespor et al. (1996) megfigyelték, hogy a fonológiai prominencia helyzete korrelál a vonzat és az őt megkövetelő mondatrész viszonylagos sorrendjével. A „vonzat elől” nyelvekben a hangsúly a bal oldalon, a „vonzat hátul” nyelvekben a jobb oldalon van (pl. *on the table, az asztal alatt*). Az elméletek egy másik csoportja (Saffran et al. 1996, Mintz et al. 2002) a nyelvi bemenet statisztikai tulajdonságainak szerepét hangsúlyozza. Bizonyos elemek gyakran fordulnak elő együtt, pl. az egy szóba tartozó hangok, vagy az egy szintaktikai egységet alkotó szavak (pl. névelő + főnév stb.).

Ezen elméleti háttér alapján a következő kérdések megválaszolására tervezünk kísérleti eljárásokat:

- Van-e pszichológiai realitása azoknak a korrelációknak, amelyeket a fonológiai csizmahúzás feltételező elméletek leírnak? Más szóval, vajon a csecsemők kihasználják-e a nyelvtanulás során azokat az empirikus együttjárásokat (pl. prominencia és szórend), amelyeket a világ nyelveinek tanulmányozása során megfigyeltek?
- Vajon a statisztikai tulajdonságok elegendő információt szolgáltatnak-e általánosabb szabályok elsajátításához is, vagy csak konkrét, a nyelvi bemenetben előforduló elemek együttjárásait, összetartozásait képesek meghatározni?

Az első kérdést a következő kísérletekkel igyekszünk megválaszolni. Nyolc hónapos csecsemőket olyan ingerekhez habituálunk, amelyek a szóban forgó fonológiai tulajdonságokkal rendelkeznek, pl. meghatározott helyzetű hangsúlyuk van, majd a fejforgatási (*head turn procedure*), illetve a preferenciális nézési eljárás (*preferential looking procedure*) segítségével meghatározzuk, hogy a csecsemőknek van-e valamilyen elvárásuk a szórendet illetően. Ha a prominencia helye és a szórend közötti összefüggés valóban csizmahúzó jelzés a nyelvtanulásban, akkor azt várjuk, hogy pl. egy bal oldali hangsúlyhoz habituált csecsemő a teszt fázisban egymás után bemutatott „vonzat elől”, illetve „vonzat hátul” szórendű szerkezetek közül a hangsúllynak megfelelő „vonzat elől” szerkezeteket válassza, ami azzal jelez, hogy hosszabb ideig figyel ezekre az ingerekre vagy többször fordítja a fejét arra az oldalra, ahonnan ezeket az ingereket hallja.

A második kérdés vizsgálatára felnőtt és gyermek kísérleteket egyaránt tervezünk végezni. Ezek lényege, hogy a kísérleti személlyel olyan mesterséges nyelvtant tanítunk meg, melynek adott statisztikai tulajdonságai vannak, például bizonyos elemek (szótagok vagy szavak) sokkal gyakrabban fordulnak elő benne, mint mások. Ha a statisztikai tulajdonságoknak valóban van szerepe, akkor azt várjuk, hogy a kísérleti személyek a gyakori elemeket a természetes nyelvek ún. funkciószavaira vagy grammatikai elemeire vetítik le (pl. névelők, névutók, kötőszavak stb.), a ritkább elemeket pedig a tartalmas vagy lexikai szavakra (főnevek, igék stb.), mivel az előbbieket lényegesen gyakoribban a természetes nyelvekben, mint az utóbbiakat. Emellett, ha a megfeleltetés megtörténik, akkor feltehetjük azt a kérdést is, hogy van-e a személyeknek elvárása a gyakori és a ritka elemek viszonylagos sorrendjével kapcsolatban, attól függően, hogy anyanyelvükben milyen szerkezetek a gyakoriak.

A magyar nyelv vonatkozásában e kísérleteknek különös jelentősége van, hiszen szórendi szempontból a magyar nem tiszta eset, mind „vonzat elől”, „mind vonzat hátul” típusú szerkezetek találhatóak benne (pl. *almát eszik vs. eszi az almát*). Éppen ezért érdekes, hogy vajon a fonológiai tulajdonságok ilyen esetben milyen információval szolgálnak a gyermek számára. Korábbi vizsgálataink (Pléh, 1998) mér feltárták a két szórendi alaptípusra vonatkoztatt értelmezési eljárásokat, most ezek eredett vizsgáljuk.

Ezenkívül a magyar agglutináló jellege is kifejezetten érdekes, hiszen statisztikai szempontból a szótó + rag kapcsolatok gyakoriak, ugyanakkor a töveknek és a ragoknak önálló előfordulási gyakoriságuk is van.

Felmerül tehát a kérdés, hogy az elsajátítás során a gyermek kialakít-e szótó + rag típusú reprezentációkat is azok kiemelt gyakorisága alapján. Ez a kérdés szorosan összefügg a felnőtt lexikai reprezentáció Lukács, Pléh és mások által vizsgált tulajdonságaival (Pléh és Lukács, 2001).

2.3 Az atipikus nyelvi fejlődés vizsgálata fejlődési zavarokban

2.3.1 Nyelvi képességek vizsgálata a nyelvi fejlődés speciális zavarában

A nyelvi fejlődés zavarainak kutatása a nemzetközi kognitív pszichológia nyelvi kutatásaiban már mintegy 20 éve az egyik legizgalmasabb és legtöbbet kutatott téma (pl. Bishop & Leonard, 2000; Fletcher & Ingham, 1995), magyar nyelvterületen azonban eddig nagyon kevés szisztematikus és elméletileg motivált kutatás történt. A kutatás során olyan nyelvi tesztek szeretnénk kidolgozni és használni, amelyek lehetővé teszik egyrészt a specifikus nyelvi zavarral élő gyerekek nyelvi problémáinak feltárását és a zavarral kapcsolatban kidolgozott versengő elméletek tesztelését magyar gyerekekkel. Általános megfigyelés, például, hogy a Specifikus Nyelvi Zavarban szenvedő gyerekeknek több nyelven is az egyik legnagyobb nehézséget a finit igemorfológia (vagy annak valamely területe) jelenti. Az SLI egyik magyarító elmélete, Clahsen (1991) szerint specifikus egyeztetési deficitet figyelhetünk meg náluk. Egy másik nagy elmélet, a „kiterjesztett opcionális infinitívuszi magyarázat” (Rice et al. 1995) szerint az egyeztetés és időjegyek egyaránt sérülnek. A magyar azért különleges tesztelési terepe ezeknek az elméleteknek, mert az egyeztetésnek és az igeidőnek is gazdag morfológiája van, ugyanazokon az igealakokon vizsgálható mindkét jelenség, továbbá az alannal való szám- és személybeli egyezéssel túl egy korábban még nem vizsgált ritkább egyezéstípust, a tárgy definitívval való egyezést is vizsgálhatjuk. Az általunk összeállított teszt tehát az igeragozást vizsgálja az egyeztetés különböző fajtái és igeidő szerint. Az ige egyezik az alannal számban és személyben, a tárggyal pedig határozottságban (alanyi-tárgyas ragozás.)

A tesztek kiválasztásában és kidolgozásában három általánosabb szempont irányít bennünket:

- a tesztelésre kiválasztott nyelvi jelenség olyan legyen, hogy a belőle nyert adatok potenciálisan a specifikus nyelvi zavar valamely elméletét támogassák vagy cáfolják, ilyen terület például az egyeztetés vagy a szabályos-kivételes ragozás, vagy a verbális munkamemória vizsgálata
- vizsgáljunk olyan jelenségeket, amelyek a magyarra specifikusak, vagy legalábbis a korábban sokat vizsgált nyelvekre nem jellemzőek: ilyen az igeikötőmozgatás vagy az argumentumszerkezet részeként megjelenő esetragok,
- a gyakorlati szakemberek számára olyan tesztek kínálni, amelyek lehetővé teszik a nyelvfejlődési zavarok különböző altípusainak specifikusabb diagnózisát és fejlesztését.

A fenti megfontolásokat figyelembe véve vizsgálatainkban a következőket szeretnénk tesztelni:

- verbális rövidtávú memória
- az ige jelölt idő és egyezés különböző fajtái: szám/személy egyeztetés az alannal; határozottságbeli egyezés a tárggyal
- az igeikötők mozgatását megkívánó konstrukciók
- argumentumszerkezet: esetjelölők vonzatokban és szabad határozókban
- szabályos és kivételes morfológia.

2.3.2 Autizmus és pragmatika

Az autizmus, illetve tágabban, az autizmus spektrum zavarok (ASD) elsősorban azért képviselnek kitüntetett kutatási területet a nyelv mentális reprezentációjának és elsajátításának megértésében, mert számos korábbi vizsgálatból (összefoglalásért l. Tager-Flusberg, 2000) úgy tűnik, e neurokognitív fejlődési zavarokban már az elsajátítás során disszociáció jelentkezik a nyelv formális (fonológiai-grammatikai-szemantikai) illetve kommunikációs-pragmatikai aspektusai között. Erős érvek sorakoztathatóak fel azon feltételezés mellett is, hogy ez utóbbi aspektusok sérülése mögött az intencionalitás megértésének, a naiv tudatelméleti mechanizmusnak a sérülése áll, legalább részben. Az autizmus (ASD) úgynevezett „magasan funkcionáló” esetei így különleges betekintést kínálhatnak abba, miként, milyen mechanizmusok által függenek össze a nyelv formális és társas-kommunikatív aspektusai az elsajátítást, majd az érett funkciót tekintve, illetve mennyiben és miként (milyen mechanizmusok által) támaszkodik az elsajátítási folyamat a naiv tudatelméleti képességre. A jelen kutatási területen ezen problémakörbe tartozó konkrét hipotéziseket vizsgálunk, alapvetően

- autizmussal élő,
- tipikusan fejlődő, illetve esetenként
- specifikus nyelvi zavarral küzdő,

különböző életkorú illesztett (életkor, IQ, nyelvi profil különböző aspektusai) minták (15-30 fő) összehasonlító vizsgálata révén, részben korrelációs, részben kísérleti designokat alkalmazva. A vizsgálatok során ugyanakkor

különös hangsúlyt helyeznénk az egyéni fejlődési variációkra, különösen a két atipikus fejlődésű csoport esetében, és elsősorban saját, korábbi vizsgálataink eredményeire, illetve az azok keretében fejlesztett vizsgálati eljárásokra támaszkodnánk (l. Győri és mtsai., 2004).

A vizsgálandó főbb kérdések a következők:

- A nyelv formális aspektusainak elsajátítása és annak kapcsolata a verbálisan illetve non-verbálisan manifesztált tudatelméleti képességgel, tipikus illetve atipikus fejlődésű (ASD és SLI) gyermekeken.
- A verbálisan illetve non-verbálisan manifesztált tudatelméleti képesség viszonya és disszociáció autizmusban.
- Munkaemlékezeti és végrehajtó funkció korlátok a nyelv és a tudatelmélet elsajátításában és működésében autizmussal élő gyermekeknél.
- Egyéni variációk a tudatelmélet és a nyelv formális aspektusainak disszociációjában autizmussal élő és specifikus nyelvi zavarral küzdő gyermekeknél.

2.3.3 Williams-szindróma

Műegyetemi kutatócsoportunk négy éve foglalkozik a Williams-szindrómások (WSz) nyelvi, emlékezeti és vizuális funkcióinak vizsgálatával. Eddigi eredményeink (Pléh et al., 2003; Lukács et al., to appear) számos további vizsgálat szükségességét vetik fel. A téri kifejezések használatában a Williams-szindrómában és tipikus fejlődésben is megfigyelt mintázatra (a forrás jellegű kifejezések különös nehézsége) egyik magyarázatunk a téri-vizuális munkamemória kis kapacitása. Ezt a hipotézist szeretnénk tesztelni olyan helyzetekben, amelyek jobban kontrollálják a munkamemória kapacitás helyzetére gyakorolt hatását. Érdemes lenne a WSz csoport teljesítményét a téri munkamemória valamilyen mutatóján illesztett kontrollcsoporttal összehasonlítani. Egy további lehetőség a Williams-szindrómások felismerési emlékezeti teljesítményét vizsgálni ugyanazokra az elrendezésekre, amelyeket a nyelvi feladatban teszteltünk, vagy arra kérni a személyeket, hogy az elrendezések bemutatása után azonnal reprodukálják forrás, statikus és cél jellegű elrendezéseket. Korábbi anaforamegértési vizsgálatunkban a WSz és a verbális korban illesztett kontrollcsoport is véletlen szinten teljesített, és ezt lehet, hogy a vizsgálati elrendezés és módszer (a (11a-b)-hez hasonló mondatok utáni kérdésekre kellett válaszolni) nehézsége okozta.

(11) a. A pingvint megállította a kacsá. Ø Felborította a szekrényt. Ki borította fel a szekrényt?

(11) b. A pingvint megállította a kacsá. Az felborította a szekrényt. Ki borította fel a szekrényt?

Ezt a feladatot szeretnénk átalakítva megismételni, az eddigi mondatértelmezéses helyzetből egy kisebb emlékezeti terhelésű képválasztásos feladattá.

3. Pszicholingvisztika és a nyelvtechnológia egymáshoz kapcsolása

3.1 A számítógépes alaktani elemzés és a tényleges emberi feldolgozás modelleinek közelítése.

Feltételezhető, hogy a gépi alaktani elemzők architektúrája és az emberi szóforma feldolgozás algoritmusai egymás modelljeiként szolgálhatnak. Számos pszicholingvisztikai kísérlet megmutatta például, hogy bizonyos esetekben morfológiailag komplex alakokat egészen távolunk a mentális lexikonban (ha pl. ha az alak gyakori vagy a jelentése nem transzparens). Ennek megfelel az a gépi hatékonyságnövelő megoldás, amikor néhány ezer gyakori szót az úgynevezett gyorsító (cache) memóriában tárolunk, hogy ezzel a szövegeket feldolgozásának sebességét növeljük. Általánosítva feltételezhetjük, hogy pontosabb pszicholingvisztikai modelleket alkothatunk és a gépi megoldások hatékonyságát növelhetjük a gépi és az emberi feldolgozás empirikus összehasonlításával.

A morfológiai elemzők fejlődésében három világos fázist különíthetünk el. Az első, és máig gyakran alkalmazott módszer a balról-jobbra haladó tő-, illetve a jobbról-balra haladó rag-levágás, illetve ezek különféle kombinációi. Ilyen a magyar HuMor, a nemzetközi Ispell (36 nyelven, köztük magyarul is) helyesírás-ellenőrző, az angol alkalmazásokban nagyon népszerű Porter tövező (Porter 1980) és a magyar nyelvre elkészült szabadon elérhető, nyitott HunMorph morfológiai elemző (Halácsy et al 2004). Ezekben az algoritmusokban a kontrolllogika és a nyelvtani szabályok még gyakran keverednek programozás-technikailag is és konceptuálisan is.

A második módszer a véges állapotú „tokenizálás”, ahol a programlogika és a nyelvtani logika világosan szétválik. Ezek a rendszerek korábban az unix közismert lex/flex programcsomagján alapultak, de Johnson (1970), Koskenniemi (1983), Kaplan & Kay 1981, Kaplan & Kay 1994) és mások munkáinak hatására ma már a gyakorlatban a morfológiai elemzést leggyakrabban kétszintű véges állapotú automatákkal (angolul finite state transducer, FST), és a gyakori többértelműség miatt ezek nemdeterminisztikus változatával szokták elvégezni.

A harmadik módszer, a statisztikai alapú elemzés is tulajdonképpen a nemdeterminisztikus állapotautomaták általánosítása valószínűségekkel súlyozott, valamint rejtett állapotterű (rejtett Markov) automatákra (Kornai 1992, Mohri 1996, Hajíč 2001, Clark 2001, Hakkani et al. 2002). A három módszer nagyban eltér a feldolgozási sebességben, a prediktív derivációk kezelésének módjában és az ismeretlen szavak tanulási képességében (talán

utóbbiról részletesebben is írhatunk). Az alaktanilag komplex szavak menetközbeni humán feldolgozásának sebességmérésére kidolgozott módszerek (pl. lexikális döntés, előfeszítési paradigmák) lehetővé teszik, hogy összevessük az emberi feldolgozás sebességét befolyásoló tényezőket (szuffixumok száma: *baráttól* vs. *barátságatlanságaimtól*, típusa (képző vs. rag): *barátság* vs. *barátot*; a paradigma mérete, a toldalékolás fonológiai és szemantikai áttetszősége).

Ahhoz, hogy gépi és az emberi feldolgozást össze tudjuk hasonlítani és a legjobban illeszkedő módszert ki tudjuk választani, első lépésben a programlogikát (algoritmust) és a nyelvtani logikát (nyelvtant) kell függetlenítenünk. A magyar morfológiának (a fent említett humorph elemzőhöz) már készült egy számítógépes leírása, amit első lépésben fel tudunk használni az affixum levágásos módszer és az emberi feldolgozás összehasonlításához. Kutatási célra felhasználható kétszintű véges állapotú morfológiai adatbázis és szabályrendszer a magyar nyelvhez még nem készült el; így második lépésben át kell fordítani az előbb említett adatbázist kétszintű elemzőhöz (Xerox xsft). Statisztikai alapú elemző fejlesztésére nem vállalkozhatunk. Ezen kutatás kereteiben, azonban korpuszelemzéssel súlyozhatjuk a kétszintű morfológiai adatbázisból készült hálózat állapotátmeneteit. Ezek alapján a harmadik lépésben becsülhető, hogy a statisztikai elemzők adott szóalakra milyen feldolgozási sebességgel működnek. Negyedik lépésben a produktív derivációk kezelésének lehetőségét hasonlítjuk össze a háromféle rendszerben, és vetjük össze az emberi feldolgozással. Az utóbbira hibadetekciós, szövegfelismerési és kiolvasási vizsgálatokat használunk.

3.2 Gépi szótárak és az emberi mentális szótár asszociatív hálózatainak kutatása

A szavak és fogalmak szerveződése hierarchikus és egyben többféle alapon (például személyes élmények, ellentétek, gyakori együttes előfordulás) szerveződő asszociációs hálózatként képzelhető el. Nagyon gyakran válik szükségessé egy kutatás kezdeti fázisában a vizsgált szavak asszociációs mintázatának a bemérése. A módszertan lényege, hogy a személyeknek a megadott szavakról elsőként eszükbe jutó szót vagy megadott időn belül esetleg több szót kell megadniuk. Kérhetünk a hívó szóra kategória nevet vagy ellentétet is. Az így kapott asszociációkat gyakoriságuk szerint értékeljük ki, majd a továbbiakban szemantikai kapcsolataik vagy szintaktikai funkciójuk szerint elemezhetjük őket.

A számítógépes nyelvészet régóta foglalkozik a szóasszociációk automatikus feltáráásával (többek közt teauruszok automatikus építéséhez, szinonimák kereséséhez stb., lásd Pierre 2003), de ezeket soha sem vetették még módszeresen össze az emberi asszociatív hálózat felépítésével. Vélhető, hogy ebből a munkából nem csak a pszicholingvisztika, hanem a természetes nyelvi feldolgozó programok is profitálnak, például a keresőprogramok hatékonyságának növelése révén (Jing 1994).

Célunk a kutatás keretében korpusz elemzés alapján különböző módszerekkel szóasszociációs térképek építése, és ezen hálózatok tulajdonságainak összehasonlítása az emberi hálózatokkal. Ezt a Morphologic-nál kidolgozott webes technológiával tudnánk vizsgálni (egy speciális website-on lehetne online kérdéseket megválaszolni, asszociált szavakat gyűjteni stb., ennek gyors megvalósításához van működő technológia.)

3.3 Számítógépes szógyakorisági adatbázis telepítése

A nyelvvel kapcsolatos kutatásokhoz elengedhetetlen a szavak előfordulási gyakoriságának az ismerete. Egyelőre spontán beszédből gyűjtött adatbázisok nincsenek, ehelyett az írott szövegeken alapulnak a létező adatállományok. A BME-n épített szószabalya webkorpusz alapján készült gyakorisági szótár, mely magyar weboldalak (így szépirodalom, sajtó, fórum, üzleti, műszaki szövegek) alapján készült. Nemcsak a korpusz, hanem a teljes korpuszpépítő és -kezelő programcsomag is hozzáférhető, melynek segítségével képesek leszünk újabb, finomabb regisztereken alapuló gyakorisági szótárak is létrehozni.

A pszicholingvisztikai kutatásokhoz azonban nem elégséges egy egyszerű szólista, hanem a morfológiailag, pontosabban a morfoszintaktikailag egyértelműsített korpusz alapján kell elkészíteni a gyakorisági szótárat. Különböző szempontok, például az egyes tövváltozatok vagy a toldalékok és toldalékkombinációk szerint szabadon kereshetővé kell tenni azt.

3.3.1 Szógyakoriság szubjektív megítélése

A kutatások azt mutatják, hogy minden reakcióidő mérő feladatot nagyban befolyásol, hogy az alkalmazott szavak milyen gyakorisággal fordulnak elő az adott nyelvben. Az írott szövegek (szépirodalom, napilapok, folyóiratok) alapján készített gyakorisági adatokat viszonylag jól tükrözi a szógyakoriság szubjektív megítélése. Az írott szövegek alapján számított gyakorisági értékek, amelyekre a kísérletekben támaszkodunk, mindazonáltal félrevezetőek olyan esetekben, ahol a hétköznapi szóhasználat jelentősen eltér az újságok nyelvezetétől. Ilyen esetekben, és általában a vizsgálati anyag kontrolljaként, fontos lehet szubjektív szógyakoriság becsléseket gyűjteni a vizsgált populációból. Az eljárás keretében a személyekkel gyakorisági skálakon ítéltetjük meg, hogy a kísérletünkben használni szándékozott szavak szubjektív megítélésük szerint milyen gyakoriak.

3.4. Szemantikai hálók embernél és a nyelvtechnológiában

Az ismert gépi szemantikai háló formalizmusok (WordNet és kiterjesztései) és a magyar nyelv kapcsolatának vizsgálatát tüzzük ki célul: mennyiben alkalmazható a WN formalizmus a magyar szavak és fogalmak leírására a különböző szófajoknál. Pl. főneveknél WN relációk rendben vannak, de sok specifikus magyar reláció lehet a termékeny derivációs morfológiai szabályok miatt. Magyar igék rendszere mennyire írható le a WN formalizmussal, a vonzatkeret-információk, igekötők stb. szerepe. Szeretnénk összehasonlítani az angol és a magyar fogalmi hierarchiákat (kétnyelvű angol-magyar WN vizsgálata, esettanulmányok a készítésről) főneveknél és igéknél (esetleg mellékneveknél és határozószóknál is). A kutatás végén egy elméleti áttekintő fejezetben áttekintենék a különböző gépi lexikon-leíró formalizmusokat és ontológiákat és célunk a WordNet reprezentáció felépítésének bemutatása a különböző szófajoknál (főnév, ige, melléknév, határozószó); a WN pszicholingvisztikai eredetének/motivációjának áttekintése; a WN-tel szemben megfogalmazott kritikák (pszicholingvisztikai, elméleti nyelvészeti, lexikográfiai és nyelvtechnológiai szempontok) ismertetése; egyéb, szabadon felhasználható szemantikai hálók, ontológiák, többnyelvű szemantikai adatbázisok rövid bemutatása is.

4. Kimenet

A kutatás célja a magyar nyelv mentális reprezentációjának, vagyis az egyes beszélők fejében történő leképezésének feltárása. Ennek segítségével kialakítunk egy sztenderd hivatkozási alapot képviselő kézikönyvet, amely a pszicholingvisztikát használók és alkalmazók számára kiindulópont lesz. Három területen illetve módszertannal folytatunk kutatásokat ennek a célnak az elérése érdekében.

Kutatásunknak háromféle eredménye lesz:

1. Konkrét empirikus és elméleti kutatási dolgozatok. Lásd a fenti tervet..
2. Az elkészült nyelvi kísérleti szoftverek magyar anyagra nyilvánosak lesznek.
3. Kézikönyv. Ennek előzetes vázlatát adjuk itt meg. A kézikönyv anyagába a szakirodalom eredményeinek összefoglalása és ez elméleti kérdések mellett beépítenék a fenti részletes kutatási tervben vázolt kutatások eredményeit is.

4.1 A PSZICHOLINGVISZTIKA MAGYAR KÉZIKÖNYVE. Könyvterv

- I. ELMÉLETI RÉSZ
Nyelvek és nyelvtanok: Egyetemes és nyelvspecifikus mozzanatok
A pszicholingvisztika története
A pszicholingvisztika nyelvészeti, kognitív és idegtudományi keretei
- II. MÓDSZERTANI RÉSZ
Kísérleti eljárások és a mondatmegértés kutatása (szófelismerés, monitorozás, döntések, szubjektív ítéletek)
Fontosabb kísérleti szoftver rendszerek
Percepció és emlékezeti eljárások a pszicholingvisztikában
Olvasási módszerek
Szemmozgás regisztrálás
Korpusz eljárások a pszicholingvisztikában
Neurobiológiai módszerek
Gyermeknyelvi kísérletek szervezése
Adatelemzés a gyermeknyelvben
- III. A MEGÉRTÉS KÍSÉRLETI VIZSGÁLATA
Architektúrák és nyelvtanok a megértésben
A szófelismerés és a mentális szótár
A magyar alaktan különlegességei
Tematikus szerepek a mondatokban
Összehasonlító perspektíva a magyar megértésre
Szemantikai és pragmatikai integráció: Mondat és szöveg
Megértés és integráció a percepcióval

- IV. A GYERMEKNYELV FEJLŐDÉSE ÉS MEGHATÁROZÓI
 Nyelvtanírás és a gyermeknyelv kibontakozása
 A nyelvtani kategóriák és szabályok: megfigyelései és kísérleti adatok
 A kognitív alapok: a téri nyelv példája
 Kognitív alapok és szociális kompetencia a kétnyelvűségben
- V. FEJLŐDÉSI ZAVAROK
 A specifikus nyelvi elmaradás elméletei és magyar adatai (SLI)
 Az autizmus és a pragmatika
 A Willims szindróma: Miben sajátos, miben nem?
- VI. A NYELVI REPREZENTÁCIÓ ÉS FELDOLGOZÁS SZERZETT ZAVARAI
 Afáziák
 Aspecifikusabb hatások
 Nyelv és készség, nyelv és tudás
- VII. NYELVTECHNOLÓGIA ÉS EMBERI NYELVFELDOLGOZÁS

5. Irodalomjegyzék

- Baddeley, Alan (2003). Working memory and language: an overview. *Journal of Communication Disorders*, 36, 189-208
- Bishop, D. & Leonard, L. (Eds.) (2000) Specific language impairment in children. Hove, UK: Psychology Press.
- Bloom, M. (2000). *How children learn the meaning of words?* Cambridge, MA: MIT Press
- Caplan, D. & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, Vol. 22, Number 1., 77-126.
- Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge U. Press
- Clahsen, H. (1991) *Child language and developmental dysphasia*. John Benjamins: Amsterdam, Philadelphia.
- Clark, Alexander (2001). Learning Morphology with Pair Hidden Markov Models. *Proceedings of the Student Workshop at ACL 2001*. Toulouse. France. 55–60.
- Fletcher, P. and R. Ingham (1995) Grammatical impairment. In Fletcher, P. and B. Macwhinney (eds.) *The Handbook of Child Language*. Oxford: Blackwell. 603–623.
- Gopnik, A. és Meltzoff, (1997): *Words, thoughts and theories*. Cambridge, MA: MIT Press
- Győri M. Várnai Zs., Gy. Stefanik K. (2004): A nyelv és a tudatteória elsajátítása: a kölcsönhatások jelentősége és természete. In: Győri M. (szerk.): *Az emberi megismerés kibontakozása: társas kogníció, emlékezet, nyelv*. Budapest: Gondolat.
- Hajjč, Jan (2001). Statistické modelování a automatická analýza přirozeného jazyka (morfologie, syntax, překlad) URL: <http://ufal.mff.cuni.cz/publications/year2001/slovko1.doc>
- Hakkani-Tür, Dilek, Oflazer, Kemal, Tür, Gökhan (2002). Statistical Morphological Disambiguation for Agglutinative Languages. *Journal of Computers and Humanities*. Volume 36. No. 4. 381–410.
- Halácsy Péter, Kornai András, Németh László, Rung András, Szakadát István, Trón Viktor Leveraging the open source ispell codebase for minority language analysis. *Proceeding of the 4th International SALT MIL (ISCA SIG) LREC workshop*
- Johnson, C. D. (1970). Formal aspects of phonological representation. Ph.D. dissertation. University of California. Berkeley.
- Kaplan, R. M – Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics*. 20(3). 331–378.
- Kaplan, R. M. – Kay, M. (1981). Phonological rules and finite-state transducers. *Linguistic Society of America Meeting Handbook. Fifty-Sixth Annual Meeting*. New York. Abstract.
- Kornai András (1992). Frequency in morphology. I. Kenesei (szerk.): *Approaches to Hungarian IV*. 246–268.
- Koskeniemi, K. (1983). Two-level Morphology: A General Computational Model for Word-form Recognition and Production. Univ. of Helsinki. Dept. of Gen. Ling. Publications No.11.
- Lukács, Á., Cs. Pléh, & M. Racsmány (megjelenés alatt): Language in Hungarian Williams syndrome children. in Niedeggen-Bartke, Susanne & Julia Siegmüller (szerk.): *Linguistic development and competence in Williams syndrome*
- MacWhinney, B. (2003). *The CILDES system*. Machlach: Erlbaum.
- Marslen-Wilson, W.D., Tyler, L.K., Waksler, S. & Older (1994) Morphology and meaning in the mental lexicon. *Psychological Review*, 101 (1), 3–33.
- Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, 26, 393-425.
- Mohri, Mehryar (1996). On some Applications of Finite-State Automata Theory to Natural Language Processing. *Journal of Natural Language Engineering*. 2. 1–20.

- Morgan, J. & K. Demuth (eds.). 1996. *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates. 487 pages.
- Nespor, M., M.-T. Guasti and A. Christophe (1996) Selecting word order: the Rhythmic Activation Principle. in: U. Kleinhenz (ed.) *Interfaces in Phonology*. Berlin. Akademie Verlag. pp. 1-26.
- Pierre P. Senellart and Vincent D. Blondel (2003). Automatic discovery of similar words. *A Comprehensive Survey of Text Mining* Springer-Verlag
- Pinker, Steven. 1991. *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pléh Cs. (1998): *A mondatmegértés a magyar nyelvben*. Budapest: Osiris
- Pléh Cs. és Lukács Á. (szerk., 2001). *A magyar morfológia pszicholingvisztikája*. Budapest: Osiris
- Pléh, Cs., Á. Lukács and M. Racsmány (2003) Morphological patterns in Hungarian children with Williams syndrome and the rule debates. *Brain and Language*, 86, 377–83.
- Porter, M. (1980). An algorithm for suffix stripping. *Program* 14(3):130–137 , 1980.
- Rice, M., Wexler, K. and P. Cleave (1995) Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research*, 38, 850–863.
- Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926-1928.
- Tager-Flusberg, H. (2000): Language and understanding minds: connections in autism. In: Baron-Cohen, S., Tager-Flusberg, H. and Cohen, D. J., eds., *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. (2nd edition) Oxford: Oxford University Press.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, Mass.: Harvard University Press.
- Y. Jing and W. Bruce Croft (1994). An association thesaurus for information retrieval. Proceedings of {RIAO}-94, *4th International Conference "Recherche d'Information Assistée par Ordinateur"*. 146–160.