

A magyar INTEX fejlesztéséről

Tamás Váradi and Gábor Kata

Kivonat

Az INTEX rendszer egy véges állapotú technológián alapuló, integrált számítógépes nyelvelemző szoftver eszköz, melyet közel egy évtizede fejlesztett ki Max Silberstein Maurice Gross elvi irányításával. Az INTEX egyaránt kiválóan alkalmazható kutatási és oktatási célokra, és a francián kívül számos nyelvre sikerrel alkalmazták. Jelen dolgozatban beszámolunk a Nyelvtudományi Intézetben folyó munkákról, melynek célja az INTEX magyar változatának a kifejlesztése.

1. Bevezetés

A dolgozatban röviden áttekintjük az MTA Nyelvtudományi Intézetben az INTEX magyar változatának kifejlesztését célzó munkálatokat. Az INTEX rendszer egy robosztus, komplex nyelvelemző keretrendszer, amely igéretes eszköznek tűnik mind a számítógépes nyelvészeti kutató-fejlesztő munka, mind a (számítógépes) nyelvészeti oktatás számára. A rendszert francia kutatók fejlesztették ki Maurice Gross irányítása alatt, közel egy évtizedes története során számos más nyelvre (angol, olasz, szerb, portugál) is alkalmazták. A véges állapotú technológiát és az elektronikus szótárakat széles körben alkalmazó rendszer magyarra ültetése nem triviális feladat. Megoldása viszont azzal kecsegtet, hogy nemcsak saját munkánk számára nyerünk egy általánosan használható eszközt, de a nem számítógépes nyelvész szakemberek kezébe is tudunk adni egy olyan szoftvert, amellyel nyelvtant, lexikai erőforrásokat, nyelvelemző rendszereket tudnak fejleszteni.

A dolgozat az alábbi részekből áll: A 2. részben ismertetjük az INTEX főbb jellemzőit. A 3. összefoglaljuk azokat a problémákat, amelyeket a magyar változat elkészítéséhez meg kell oldani. A 4. és 5. rész bemutatja a magyar rendszerrel végzett eddigi munkát a nyílt tokenosztályok felismerése illetve a felszíni szintaktikai elemzésében.

2. Az INTEX rövid jellemzése

Az INTEX számítógépes nyelvészeti fejlesztő rendszert a Párizs VII Egyetem LADL laboratóriumában Max Silberstein készítette [5] és azóta is folyamato-

*MTA Nyelvtudományi Intézet, 1068 Budapest Benczúr u 33, e-mail: {varadi,gkata}@nytud.hu

san fejleszti. A szoftver szellemi atyja Maurice Gross, a laboratórium vezetője, akinek a lexikonra épülő, bottom-up nyelvelírási modellje [3] egyértelműen inspirálta a rendszer elvi felépítését. A rendszer a fenti történeti okokból a francia nyelvre van leggazdagabban kidolgozva, de az angol mellett az olasz, bolgár, portugál, spanyol változatai is többé-kevésbé léteznek.¹ Magyarországon maga a fejlesztő ismertette a COMPLEX'99 konferencián, de érdemi munkálatok a magyarra csak az utóbbi két évben kezdődtek.

A rendszer meghonosítását nemcsak a robosztus és gyors véges állapotú technológia indokolja, hanem a fejlesztőknek az a kifejezett szándékuk, hogy egy viszonylag könnyen használható oktatási eszközt is adjanak a nem informatikus nyelvészek számára. Első megközelítésben az INTEX egy gyors korpuszkezelő eszköznek tűnik, amely amint betöltöttünk egy sima ascii szöveget, máris készen áll arra, hogy lekérdezhessük reguláris kifejezések segítségével. A reguláris kifejezések azonban nemcsak a szavak alakjára hanem nyelvi (morfoszintaktikai vagy akár szemantikai) jegyeikre is utalhatnak. Ezek az információk a szótári komponensből származnak, amely a rendszer központi részét képezi.

A szótár egy- illetve többtagú kifejezések tára, melyekben szóalakok találhatóak, a lemmával és tetszőleges társított nyelvi információval, mindez igen hatékony véges állapotú belső reprezentációban. A rendszer egyedi sajátossága, hogy a szótár, a szöveg valamint a szövegre alkalmazott grammatika mind egyaránt véges állapotú technológiával van megvalósítva. Ami a rendszert széles körben is különösen használhatóvá teszi az a grafikai felület, amelyen viszonylag egyszerűen szerkeszthetjük és kezelhetjük a véges állapotú tranzducereket (ld. példaként a 1. ábrát).

A szöveg betöltése után előfeldolgozó, normalizáló grammatikákat futtathatunk rajta, majd kiválaszthatjuk az alkalmazandó lexikai erőforrásokat. A szótárak lefuttatása azt jelenti, hogy a szöveg minden szavához társul a szótárban tárolt információ, melyeknek minden elemére külön-külön is hivatkozhatunk a további feldolgozás során. Óhatatlanul lesznek természetesen ismeretlen szavak, amelyek listáját azonnal megtekinthetjük, illetve számolnunk kell többértelmű tokenekkel. Ez utóbbiak megtekintésére igen érdekes lehetőség a szöveg mondatainak átalakítás egy véges állapotú tranzducerré, melynek a kimenetét az egyes tokenek elemzése adja, és annyi bejárési utat találunk a tranzducerben, ahány féleképpen többértelmű az illető mondat. Az egyértelműsítés egyik eszköze a lexikai szűrés (a szótárak alkalmazási sorrendjének a megváltoztatásával) illetve az egyértelműsítő gráfok, azaz olyan tranzducerek, melyeknek kimenete az adott kontextusban érvényes elemzést adja.

3. A magyar változat nehézségei

A fenti vázlatos ismertetésből leszűrhető, hogy az INTEX a szótárak segítségével azonnal előállítja a szöveg morfológiai elemzését is. Ez a magyarban köztudomásúlag cseppet sem triviális feladat. Különösen azért nem, mert a francia változat azt a megoldást alkalmazza, hogy az egy-egy lemmához tartozó összes képzett és

¹Ld. bővebben az INTEX webhelyet: <http://www.nyu.edu/pages/linguistics/intex/>

ragozott alakot tétélesen felsorolja a szótárban. A „felsorolást” nem kézzel kell végezni, mert szintén gráfok segítségével definiálhatjuk az egyes paradigmához tartozó alakok szerkezetét, majd a gráfok alkalmazásával előlíthatjuk az összes alakot, melyeket végül aztán bináris alakban tömöríthetünk. Akármennyire is automatizált a folyamat, a magyar morfológia gazdagsága és produktivitása nem teszi lehetővé, hogy az Értelmező Kéziszótárban szereplő összes szó valamennyi lehetséges toldalékolt alakját előállítsuk.

Az INTEX legutóbbi változata [6] már lehetővé teszi, hogy igazi lexikai transzducereket használjunk, azaz a szóalakok felépítését gráfokkal definiáljuk, ahol pl. a töveket tetszőleges hosszú sztringként határozzuk meg. Sajnos azonban kísérleteink Elekfi László ragozási szótárából [2] készített morfológiai adatbázisunk INTEX lexikai transzducerekké alakítására nem vezettek eredményre: a legenerált toldalékszekvencia halmazok implementálása meghaladta az INTEX lexikai transzducerek optimalizálási képességeit (ld. részletesebben [8]).

Az INTEX magyar változatáról tehát akkor beszélhetünk, ha rendelkezésünkre állnak az INTEX szótárak olyan változatai, amelyek kellő számú lemmával rendelkeznek (ideálisan legalább az ÉKSz. mintegy hetvezernyi címszavával), tartalmazzák a morfológiai jellemzőket és alkalmasak az egyes címszavakhoz tartozó valamennyi szóalak felismerésére. Ebben az esetben válik ugyanis az INTEX tetszőleges magyar szöveg számára is használhatóvá.

Természetesen addig is, amíg a jelenleg folyó intenzív munkálatok végleges eredményre nem vezetnek, az a lehetőség már most is adott, hogy meghatározott korpusz számára, az abban előforduló valamennyi szóalakot elemezzünk külső eszközzel, például a HUMOR-ral [4], majd az elemzés kimenetét INTEX DELAF szótárfile-okba építsük be, és ezeket kompiláljuk. Nyilvánvaló, hogy ez csak korlátozott megoldás, hiszen csak arra a korpuszra teszi alkalmassá az INTEX használatát, melynek szókincséből a szótárat előzetesen összeállítottuk. Addot feladatok elvégzéséhez azonban ez a testreszabott megoldás is célszerűnek bizonyult: az INTEX-et sikeresen használtuk eddig is különböző nyelvtechnológiai projekteknél. A továbbiakban az ilyen, egyedileg kipreparált szótárakra épülő munkálatokról számolunk be. (A magyar morfológia INTEX-ben történő implementálásáról ld. még Nagy Viktor és Vajda Péter tanulmányát a jelen kötetben.)

4. Tulajdonnévfelismerés az INTEX segítségével

A munka első fázisaként az alábbi tulajdonnév-típusokkal foglalkoztunk: személynevek, helységnevek, illetve intézmények nevei.² Abból az alapfeltevésekből indultunk ki, hogy ezek a - többnyire több szóból álló - kifejezések általában tartalmaznak egy vagy több, nyílt tokenosztályhoz tartozó szót, valamint legalább egy úgynevezett *horgonyt*, vagyis olyan, zárt halmazhoz tartozó szót, amely segít a tulajdonnév típusának meghatározásában. Ilyen horgony például a személy-

²A tulajdonnévfelismerő rendszer Intexes változatának kidolgozásakor Sass Bálint korábbi munkáját vettük alapul. Az adatok gyűjtését és a reguláris kifejezések szerkesztését elsősorban Sass Bálint végezte Chinchor által definiált alapelvek [1] felhasználásával.

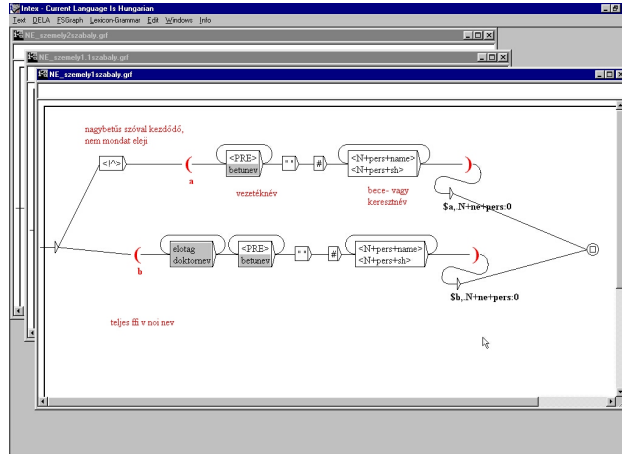
neveknél a *Dr.* vagy *Ifj.* előtag, vagy a helységneveknél az *utca, tér* utótag. Az általunk alkalmazott módszer a tulajdonneves szerkezetek e két összetevőjének kezelésére épül. A tulajdonneves szerkezetek felismeréséhez és kategorizálásához az Intex szótárait, illetve szótári lookup-programját, valamint szó-szintű és szintaktikai transzducereket használtunk. A szövegelemzéshez többféle szótár áll rendelkezésünkre: az egyszerű szavakat tartalmazó DELAF szótár, a többszavas kifejezéseket tartalmazó DELACF szótár, valamint mindkét típusból a reguláris kifejezéseket is tartalmazó, nyílt tokenosztályok leírásához használt szótár-transzducerek. A tulajdonnévfelismeréshez valamennyi típusú szótárra szükségünk volt, valamint felhasználtunk egy kézzel összeállított, többszavas szótárban kódolt kész tulajdonnév-listát, melyet a munkánk során ma is folyamatosan bővítünk.

A különböző típusú tulajdonnevek részét képező, zárt halmazhoz tartozó horgonyokról listákat készítettünk, majd ezeket a horgonyokat a Delaf (egyszavas) és a Delacf ("compound words", többszavas) szótárakban szemantikai jegyekkel láttuk el, melyek a legtöbb esetben egyértelműen azonosítják annak a tulajdonneves szerkezetnek a típusát, melyben az adott horgony előfordulhat.³ Ezzel párhuzamosan kezeljük azokat az egyszavas tulajdonneveket, melyek nyílt tokenosztályhoz tartoznak (ilyen például a -né végű asszonynév). Ezek felismeréséhez olyan lexikai transzducereket alkalmazunk, melyek szó-szintű reguláris kifejezésekkel írják le a tulajdonnév szerkezetét. A horgonyok felismerése után a következő lépésben reguláris nyelvtanokkal meg kell találnunk azokat a szavakat, melyek a horgony körül még a tulajdonneves szerkezethez tartoznak. Az Intex erre olyan eszközt kínál, mely a szintaktikai elemzést végző transzducerekhez hasonlóan működik, ám a kimenete - a szótári lookup programhoz hasonlóan - nem közvetlenül a szövegfájlba, hanem a szöveg szókincsét tartalmazó szótár-fájlok egyikébe kerül. A nyelvtani elemzés későbbi lépései számára ez azért előnyös, mert így a felismert tulajdonneves szerkezet egy szótári egységként kezelhetjük. További előnye ennek a megoldásnak, hogy mivel a tulajdonnév-nyelvtanokkal felismert sztringeket egy fájlban tároljuk, a fájl kézi ellenőrzése után a helyesnek talált tulajdonnevek hozzáadhatók a Delacf szótárhoz, melyek a kézzel előállított tulajdonnévlistát tartalmazzák. A folyamat utolsó lépéseként az előállt tulajdonnévlista alapján annotáljuk a szöveget, vagyis a felismert sztringeket főnévi csoportként címkézzük, megjelölve tulajdonnév-státuszukat és az alkategóriát, amelyhez tartoznak.

5. Szintaktikai elemzés

Munkánk célja egy olyan nyelvtani eszközlánc létrehozása az Intex korpuszfeldolgozó rendszerben, mely képes magyar mondatokból álló szövegek nyelvtani elemzésére, a tokenizálástól a mondatsegmentáláson, morfológiai elemzésen, egyértelműsítésen és a tulajdonnévfelismerésen át egészen a részleges szintaktikai elemzésig. A következőkben a szintaktikai elemzés menetét fogom bemutatni.

³Ez azonban nem mindig van így, hiszen például a személynevek szerepelhetnek közterületek vagy intézmények nevében is.



1. ábra. INTEX gráf a személynevek felismerésére

Részleges szintaktikai elemzés alatt egymásra épülő lokális nyelvtanokkal megvalósított elemzést értek, melynek nem az a célja, hogy a mondat teljes elemzési fááját felépítse, hanem bizonyos, lokálisan megragadható nyelvtani szerkezetek és ezek közötti dependencia-viszonyok feltérképezésére íródott. A nyelvtan képes feltárni az alapvető szintaktikai viszonyokat akkor is, ha a bemenetnek van olyan része, amit nem ismer fel. A mondat központi elemének az állítmányt képviselő finit igét tekintem, az összetevők dependencia-viszonyai alatt pedig elsősorban az igei vonzatkeretbe tartozó elemeknek és az ige szabad határozóinak felismerését értem. A módszer arra az előfeltevésre épül, hogy az igei állítmány teljesen felépített frázisokkal - illetve ezek fejével - lép régens-argumentum viszonyba. A fent megfogalmazott feladat eszerint az alábbi részfeladatokra osztható:

1. a frázisok megtalálása és címkézése
2. a tagmondat-határ megtalálása
3. az igei argumentumszerkezet azonosítása
4. az igehez tartozó szabad határozók, tagadósók, illetve a mondat-határozók megtalálása

(A fenti sorrend megfelel a nyelvtanok alkalmazási sorrendjének, kivéve a 4)-t, melynek egyes elemei (igekötő, tagadás felismerése) a 2) nyelvtanok után következnek.)

Az 1-4) feladatokat végző nyelvtani szabályok egyaránt használják egymás kimenetét és az alsóbb szintű nyelvi elemzések információit. A feladatokat Intex gráfokkal ábrázolt reguláris nyelvtanok látják el, melyek kimenete a szövegbe illesztett annotáció - erre hivatkoznak az elemzés későbbi lépései. Ez a módszer bizonyos szempontból ugyan rugalmatlanná teszi a nyelvtant (megnehezíti

a szabályok sorrendjének módosítását), ám ezt a hátrányt ellensúlyozza az az előny, hogy szabályaink a szintaktikai elemzés minden pontján hivatkozhatnak bármilyen típusú információra. Így kisebb annak az esélye, hogy a mondattani elemzés egy alacsonyabb szintjén fel nem ismert sztring az egész mondatot elemzhetetlenné teszi, mivel egy későbbi lokális nyelvtannak még így is helyes környezeteként szerepelhet. A megközelítés másik előnye, hogy bármelyik lépésben módosíthatjuk a korábban lefutott szabályok kimenetét. A szabályokat az Intex grafikus felületét (a nyelvtanírásra szolgáló gráf-editort) használva készítettük és teszteltük. A nyelvtanok futtatása azonban parancssorból is végezhető, így egy olyan batch-fájlt készítettünk, mely az Intex mintaillesztő és szótári lookup programjait futtatja a paraméterként megadott nyelvtanokkal és szótárakkal, általunk definiált sorrendben. Az elemzés végeredménye egy txt fájl, mely az összes annotációt tartalmazza, de a hatékonyabb tesztelés érdekében az elemzés minden lépésének kimenetét külön fájlban tároljuk. A szabályok megalkotásához és teszteléséhez egy MTI üzleti rövidhírekből álló, 231 ezer szavas korpuszt használtuk.

5.1. A szintaktikai elemzés menete

A fent leírt négy részfeladatból jelenleg az 1-2) -re rendelkezünk részletesen kidolgozott nyelvtanokkal, így a továbbiakban ezeket mutatjuk be. Az 1) feladat, azaz a kötött szórendű frázisok megtalálása és címkézése magában foglalja a főnévi, melléknévi és a névutós csoportok azonosítását. Ezek közül a főnévi csoportnak nemcsak a határait, hanem a fejét, valamint annak lényeges morfoszintaktikai és szemantikai tulajdonságait is megadjuk.

Az 1) feladatot öt különböző szintaktikai gráf végzi. Ezek közül a legfontosabb, az NP felismerését végző nyelvtan alapját az osztályon az NKFP 2/017/2001 projektum keretei közt készült, főnévi csoportok annotálására szolgáló nyelvtan kibővített, tesztelt változata szolgáltatja [7]. Az első NP nyelvtan az összes olyan főnévi csoport lefedésére készült, melyekben a főnévi fej nincs koordinálva. Természetesen a főnévi csoportot leíró gráf tartalmaz önrekurziót, hiszen a főnévi fej (melléknévi, igenévi, birtokosi) bővítményei bővíthetők további főnévi csoportokkal. Ezt egy beágyazott algráffal oldottuk meg, mely az eredeti NP-gráffal mindenben megegyezik, ám nincs saját kimenete - így a beágyazott NP-k nem kapnak saját annotációt. A következő nyelvtannak az a feladata, hogy az előző lépésben felismert főnévi csoportok fejét megjelölje és ellássa azokkal a szintaktikai és szemantikai jegyekkel, melyek az elemzés során később relevánsak lehetnek. Ezeket a jegyeket a szótár tartalmazza a megfelelő főnévi lemmához társítva (kivéve a főnév esetét kódoló jegyet, mely a morfológiai elemzés kimenete, de szintén a Delaf szótárba kerül). Az NP esetének jelölését a következő, a mellérendelt NP-k felismeréséért felelős szabály is használja, hiszen csak azonos esetben álló főnévi csoportok koordinálhatók. A koordinációs szabály egyszerűen felülírja az első NP-szabály kimenetét: ahol két, mellérendelő kötőszóval összekapcsolt főnévi csoportot talál, melyek azonos esetben állnak, ott a köztük lévő frázishatárokat eltörli. Az 1) feladat magában foglalja a főnévi csoporton kívül a mellékneves és névutós frázisok megtalálását és címkézését

is. Ezek a kötött szórendű frázisok alkotják a szöveg első szintű szintaktikai annotációját, és egyben a tagmondathatár-kijelölő nyelvtanok bemenetét.

Nyelvtanunk elsősorban az ige által vezérelt dependencia-viszonyok, azaz az igei vonzatok és szabad határozók feltérképezésére készült. Ez azzal a következménnyel jár, hogy a finit ige nélküli (tag)mondatokban nem azonosítunk állítmányi szerepű összetevőt. Állítmánynak mindig a finit igtét tekintjük, és az állítmány kijelölése megelőzi a tagmondathatárok kijelölését. Az a feltevésünk, hogy bár nem minden tagmondat tartalmaz finit igtét, a finit igeik mind külön tagmondathoz tartoznak. Így a tagmondathatár megtalálását célzó szabályok hatékonyságát azon mérhetjük, hogy sikerül-e minden két, azonos mondatban lévő finit ige között határt találnunk. A tagmondathatárra azért van szükségünk, mert feltételezzük, hogy azok az összetevők, melyek az igével dependencia-viszonyban állnak, azonos mondatban vannak vele (kivéve a tagmondat kategóriájú vonzatokat és bővítményeket).

Mivel a tagmondat szerkezetét a szabad szórend és a gyakori szerkezeti homonímia miatt rendkívül nehéz lenne leírni, ezért nem vállalkoztunk arra, hogy olyan reguláris nyelvtant írjunk, amely bemenetének két végén a két finit ige áll, és a köztük lévő összes pontot megvizsgálva dönti el, hol lehet köztük a határ. Ehelyett egy hatlépcsős lokális nyelvtant építettünk, mely azokat a horgonyokat (kötőszók, egyes névmások, központosítás) veszi alapul, melyek közül - feltételezésünk szerint - valamelyik minden határon jelen van. Ehhez megvizsgáltuk a kötőszók, kérdőszók, vonatkozó névmások és a központosítási jelek, azaz a lehetséges "horgonyok" disztribúcióját, és olyan nyelvtant építettünk, mely a horgonyok jelenlétén kívül a szöveg morfológiai és alapszintű szintaktikai elemzésére támaszkodik, azaz figyelembe veszi például a frázishatárokat. Így a már felismert frázisokon belülre nem kerülhet tagmondathatár. A hatlépcsős nyelvtant kiegészítettük egy guesser-szerű szabállyal. Erre azért van szükség, hogy ellenőrizzük az előző szabályok működését. Azokat az eseteket kell újra megvizsgálni, amikor a fentiek alapján feltételezzük, hogy a két ige között van határ, ám a korábbi szabályokkal nem találtuk meg: ilyenkor a kérdéses szövegrészben minden kötőszót, kérdőszót, illetve tagmondatok elválasztására alkalmas központosítási jelet megjelölünk lehetséges határként.

Az alábbi példa illusztrálja az elemzés kimenetét az 1-2) feladatok végrehajtása után:

[**clauseboundary**] [**np** A fogyasztói árak <**head=árak case=nom**> **np**] [**np** júniusban <**head=júniusban case=ine**> **np**] [**np** 0,3 százalékkal <**head=százalékkal case=ins**> **np**] [**predicate** nőttek **predicate**] [**Postp** [**np** májushoz <**head=májushoz case=all**> **np**] képest **Postp**], [**clauseboundary**] és [**np** 10,5 százalékkal <**head=százalékkal case=ins**> **np**] [**predicate** haladták meg **predicate**] [**np** a 2000 júniusi szintet <**head=szintet case=acc**> **np**], [**clauseboundary?**] [**np** az első fél évben <**head=évben case=ine**> **np**] [**np** az árak <**head=árak case=nom**> **np**] átlagosan [**np** 10,4 százalékkal <**head=százalékkal case=ins**> **np**] [**predicate** voltak **predicate**] [**AdjP** magasabbak **AdjP**], [**clauseboundary**] mint [**np** egy évvel <**head=évvel case=ins**> **np**] korábban [**clauseboundary**] - [**predicate** közölte **predicate**] [**np** a Központi Statisztikai Hivatal <**head=Hivatal ca**

se=nom> np] [np szerdán <head=szerdán case=sup> np].

6. Összegzés és további teendők

A dolgozatban igen vázlatosan áttekintettük az INTEX rendszert, amely igen alkalmas szoftvereszköznek ígérkezik kutatási és oktatási célokra egyaránt. Magyarra alkalmazásának elsőszámú feltétele a nyitott szótári rendszer megalkotása, ami egyben a szövegek azonnali morfológiai elemzését is biztosítja. E cél érdekében jelenleg is intenzív munka folyik együttműködésben Max Silberzteinrel, a rendszer fejlesztőjével. Időközben készül az INTEX teljesen átdolgozott változata, a NOOJ, amely várhatóan technikailag is megfelel a magyar morfológia kihívásainak.

Hivatkozások

- [1] N. Chinchor, E. Brown, L. Ferro, and Robinson P. Named entity recognition task definition. Technical report, MITRE, 1999.
- [2] László Elekfi. *Magyar Ragozási Szótár*. MTA Nyelvtudományi Intézet, 1997.
- [3] Maurice Gross. The construction of local grammars. In E. Roche and Y. Schabes, editors, *Finite State Language Processing*, pages 329–352. The MIT Press, Cambridge, Mass., 1997.
- [4] Gábor Prószéky and László Tihanyi. Humor – a Morphological System for Corpus Analysis. In *Proceedings of the first TELRI Seminar in Tihany*, pages 149–158, Budapest, 1996.
- [5] M. Silberztein. *Dictionnaires électroniques et analyse automatique de textes: Le système Intex*. Masson, Paris, 1993.
- [6] Max Silberztein. *INTEX Manual*. Université de Franche-Comté, <http://www.nyu.edu/pages/linguistics/intex/>, 2004. Translated by Michael Long, Université de Moncton.
- [7] T. Váradi. Főnévi csoportok annotálásaCLaRK rendszerben. In: *Alexin Zoltán - Csendes Dóra (szerk.): A Magyar Számítógépes Nyelvészeti Konferencia 2003 rendezvényen elhangzott előadások kötete, Szegedi Tudományegyetem Nyomdája*, pages 65–71, 2003.
- [8] Tamás Váradi. On developing the hungarian intex system. In *Proceedings of the 7th INTEX conference*, Tours, 2005.