

## Többszavas kifejezések kezelése MT szótárban

Váradi Tamás

MTA Nyelvtudományi Intézet  
1068 Budapest Benczúr u 33  
varadi@nytud.hu

**Kivonat:** A dolgozat a számítógépes alkalmazásokban, elsősorban gépi fordítórendszerekben használt szótárak felépítésének elveit vizsgálja. Az egyedi szavak megfeleltetése helyett a többszavas kifejezések minél nagyobb számú használata lényegesen csökkenti a többértelműséget. A többszavas kifejezések egy folytonos skála mentén helyezhetők el, melynek egyik végpontján a teljes mértékben rögzített szerkezetek, a másikon pedig a teljesen nyílt, azaz minden elemükben megváltoztatható szerkezetek vannak. A dolgozat bemutatja azt, hogyan lehet lokális grammatikákkal nyitott lexikai osztályt tartalmazó kifejezéseket kezelni egy kétnyelvű gépi fordítórendszer szótári komponensében.

### 1. Bevezetés

A legtöbb számítógépes alkalmazás használhatóságának kulcskérdése a szótári komponens. Robosztus számítógépes nyelvészeti lexikon építése nem képzelhető el úgy, hogy egyszerűen átültetjük a hagyományos szótár tartalmát elektronikus formára. Hagyományos szótáron olyan szótárat értünk, amelyik közvetlen „emberi fogyasztásra” készült, függetlenül attól, hogy könyv alakban vagy elektronikus adathordozón jelent meg. Amint látni fogjuk, az olvasó szerepének feltételezése alapvetően megszabja a szótár tartalmát és tálalásának módját.

A jelen dolgozatban a számítógépes alkalmazásokhoz készülő kétnyelvű szótár szerkezetét tárgyaljuk. A gépi fordító rendszerek és egyéb kétnyelvű számítógépes alkalmazások lexikai megfeleléseinek kidolgozásában alapvető probléma a többértelműség feloldása. Valamirevaló szótárban egy-egy szóhoz általában számos célnyelvi megfelelést társítanak, amelyek az adott szó különböző jelentéseinek felelnek meg. Néha azonos jelentést egy szinonima halmazzal értelmeznek, hasonlóan ahhoz, ahogy azt a Wordnet-ben találjuk. A szótári jelentések tipikusan kontextus nélküli megfeleltetéseket tartalmaznak. Annak eldöntése, hogy egy bizonyos nyelvi kontextusban az adott szó melyik jelentésével szerepel, az ún. WSD (*word sense disambiguation*) igen nehéz feladat, amelyik online alkalmazásban roosztus méretekben még nem vethető be.

Jelentősen könnyíthetjük a kétnyelvi megfeleltetések problémáját, ha nem egyes szavak, hanem többszavas kifejezések megfelelőit keressük. A többtagú kifejezések ugyanis gyakran mintegy magukban foglalják a többértelműsítésükhöz szükséges kontextust is. A „fogás” szó többek között jelentheti egy étkezés részét illetve szorítást, angolul ennek megfelelően a 'course' illetve a 'grip' főnevekkel fordíthatjuk. Kontextuson kívül a „fogás” szó inehérenszen többértelmű, viszont a „szoros fogás” kifejezésről önmagában véve is egyértelműen eldönthető, hogy a szorítás értelmezésről van szó.

## 2. A hagyományos szótárak korlátai

Kézenfekvőnek mutatkozik, hogy a számítógépes alkalmazások számára felhasználjuk a hagyományos lexikográfia eredményeit. Az embereknek szánt kétnyelvű szótárak azonban minden gazdagságuk ellenére súlyos fogyatékoságokat mutatnak. Amint látni fogjuk, ezek a hiányosságok nem egyes szótárak vagy szótárírók tökéletlenségét, vagy hanyagságát mutatják, hanem a szótárak rendeltetéséből fakadó elvi korlátok. A hagyományok szótárak rendeltetésének kitűnő összefoglalását adja Bolinger [1] alábbi meghatározása:

Dictionaries do not exist to define, but to help people grasp meanings, and for this purpose their main task is to supply a series of hints and associations that will relate the unknown to something known.

The dictionary has done its job when it gives the reader a handhold in his own experience — a pair of synonyms, a diagram, a context, a comparison, tied to any convenient reference post.

Nézzük röviden, melyek azok a jelenségek, amelyeket a hagyományos lexikográfia a fenti elvet követve egyszerűen kiiktat a szótárak érdeklődési köréből.

### 2.1 Hiányzó lexikai egység

A szótárak címszavai egyáltalán nem, vagy csak szórványosan tartalmaznak tulajdonneveket és egyéb enciklopédikus tudást hordozó lexikai elemeket, amelyek azonban nagy gyakorisággal fordulnak elő válogatatlan szövegekben, különösen hírekben. Részben az ilyen elemek pótlására szolgál az ún. nyílt tokenosztályú kifejezések felismerésére kifejlesztett technológia.

De nemcsak a szótári egységek hiányoznak, a meglévő szavak feldolgozása is nagy kívánnivalót. A gazdaságosság jegyében ugyanis a szótárírók nem tüntetik fel kimerítően az egyes címszavakból képzés vagy összetétel által előállítható valamenyi alakot. Az "automatikusan", azaz szabályosan előállítható alakokat általában mellőzik, feltételezve, hogy az anyanyelvi kompetenciával bíró szótárolvasó ezeket szükség esetén mind elő tudja állítani.

További korlát a hagyományos szótárak teljességében az az elv, hogy a lexikográfia a nyelvhasználat „időtálló” elemeinek rögzítését tűzi ki célul. Egy-egy lexikai egység szótárba felvétele előtt a szótáríró intuitív alapon, vagy újabban egyre gyakrabban korpuszgyakorisági adatok alapján, mérlegeli, hogy az adott lexikai elem „megérett-e” arra, hogy bekerüljön a szótárba. A számítógépes alkalmazások azonban

általában a nyers szövegekkel szembesülnek, melyekben jócskán fordulnak elő egyedi, esetleg kérészetű kifejezések. A nyelvfeldolgozó programok nem engedhetik meg annak mérlegelését, hogy létezőnek ismernek-e el bizonyos szavakat, kifejezéseket.

Ez utóbbi probléma elvi jellegű és feloldhatatlan, hiszen itt az élő, folytonosan változó nyelvhasználat és az azt valamilyen szempont (érték vagy gyakoriság) alapján rögzíteni kívánó norma viszonyáról van szó, amely más fogalmakkal a type/token, vagy a langue/parole, competence/performance ismert dilemmáit veti fel.

### 2.1.1 Hiányos vagy homályos (fuzzy) információ

Gyakran a hiányolható információ nem maga a lexikai elem, hanem azzal van a gond, hogy definíciójának terjedelme nem határozható meg egyértelműen: vagy teljesen hiányzik vagy homályos annak meghatározása, hogy az adott lexikai egység a jelenségek mely körére alkalmazható.

**grater** a kitchen UTENSIL (= a tool) with a rough surface, used for grating food into very small pieces: a **cheese/nutmeg grater** (OALD7) [10]

A **grater** is a kitchen tool which has a rough surface that you use for cutting food into small pieces (COBUILD)[7]

**grater** a tool used for grating food: a **cheese grater** (LDCE)[8]

A fenti három, élvonalbeli angol értelmező szótárakból vett idézetből az még talán kitalálható, hogy ételreszelőről van szó, de az már korántsem egyértelmű, hogy a két említett reszelőfajtán kívül milyen további reszelők vannak? Különösen kétnyelvi kontextusban, ahol a szótár olvasó nem tagja a célnyelvi (kulináris) kulturális közösségnek, nem tételezhető fel az (étel)reszelő használati körének egyértelmű ismerete. Számítógépes alkalmazásokban pedig jelenleg reménytelen ilyesfajta mindennapi tudásra apellálni.

## 2.2 A többszavas kifejezések típusai

Miután beláttuk annak fontosságát, hogy a számítógépes alkalmazásba épített szótár a lehető legexplicittebb legyen, és láttuk a hagyományos szótárak elvi és gyakorlati korlátait e tekintetben, vizsgáljuk meg a többszavas kifejezések néhány fontos jellemzőit.

Az 1. ábra mindegyik példájában nemzetiségneveket találunk. Az a) alatti példák olyan kifejezéseket tartalmaznak, amelyek maximálisan kötöttek, egyik elemük sem változtatható, jelentésük nem teljesen kompozicionális (legalábbis a szinkronia síkján). A spanyol nátha mellett nincs francia vagy román nátha. A többszavas kifejezéseket tárgyaló szakirodalom elsősorban ilyen rögzült kifejezésekkel foglalkozik lásd pl. [4] magyarra [3]. Ezzel szemben a c) alatti kifejezések mindkét eleme tetszőlegesen változtatható: itt a névszói kifejezés főtagját változtattuk, de a „francia iskola-rendszer” helyett vehettünk volna bármilyen tetszőleges nemzetnevet. Ezeknek a kifejezéseknek teljesen áttetsző a szintaktikai és a szemantikai szerkezetük, kompozicionalitásuk mindkét tekintetben maximális.

A kétnyelvű gépi alkalmazások számára mindkét típusú kifejezés kezelése viszonylag egyszerű esetet jelent. Az a) típusú, teljesen kötött kifejezéseket a szótár szócikkei között fel kell sorolni a hozzá tartozó egy vagy többszavas megfelelővel együtt. Itt legfeljebb csak a minél nagyobb lefedettséghez szükséges ráfordítás jelenthet gondot. A c) típusú, teljesen nyitott kifejezések viszont nem igényelnek szótári bejegyzést, szó szerint fordíthatók.

a)	<i>English breakfast</i> <i>French fries</i> <i>German measles</i>
b)	<i>English-speaking population</i> <i>French-speaking clients</i> <i>Spanish-speaking students</i>
c)	<i>French schooling system</i> <i>French wines</i> <i>French football</i>

**1. ábra** Többszavas kifejezések a kompozicionalitásuk foka szerint

A jelen dolgozatban a 1b) típusú példákra hívjuk fel a figyelmet. Az itt található kifejezések elemei egyértelműen szerkezetet alkotnak. Figyeljük meg, ha elhagyjuk a jelzőt vagy rosszul képzett kifejezést kapunk vagy lényegesen módosul a jelentés *\*speaking population*. A kifejezések jelentése sem kompozicionális, hiszen a *speaking* 'beszélő' szó nem beszédtevékenységre, hanem nyelvi kompetenciára vonatkozik. Ugyanakkor a kifejezés elemei nem olyan mértékben rögzítettek, mint azt az a) típusú példánál láttuk: a példákban szereplő jelzőket tetszőlegesen kombinálhatjuk a többi kifejezés főtagjával (*French speaking population*, *Spanish speaking students* stb.). Azt találjuk tehát, hogy a b) csoportban a kifejezések részben kötöttek, részben nyitottak. Felmerül tehát a kérdés, hogyan kezelhetjük őket egy számítógépes alkalmazásban?

Először az a kérdés merülhet fel, hogy egyáltalán kell-e velük külön foglalkozni, azaz nem járhatunk-e úgy, mint akár az a) vagy a c) csoport tagjaival. Listába foglalni őket és a célnyelvi megfelelőket az egyes elemekhez hozzárendelni, ha egyáltalán lehetséges, rendkívül veszteséges eljárás, hiszen ez azt jelentené, hogy a kombinálható elemek teljes permutációját elő kellene állítani, a megfelelőekkel együtt. Ugyanakkor a szó szerinti fordítás sem kivihető, mivel ugyan ebben az esetben történetesen létezik szó szerinti megfelelés *angolul beszélő népesség* stb., annak jelentése kétértelmű (tevékenységet és kompetenciát egyaránt takar.) Más esetben azonban, mint

például a 2a és 2b példákban, sem listába foglalni, sem alkotó elemenként szó szerint fordítani nem tudjuk őket.

2a) a twelve year old boy

2b) egy tizenkét éves fiú

Ha viszont a megfelelés nem áttetsző a két nyelv között, és ennek következtében fel kell vennünk a kétnyelvű szótárban, akkor a gépi alkalmazás számára nem folya-  
modhatunk a „példálódzás” módszeréhez, ami alkalmanként tökéletesen kielégítő a hagyományos szótárak esetében. A hagyományos szótáraknál elégséges megadni egy megfelelést a maga felszíni alakjában, és a minta kivonását rábízni a szótárolvasó (nyelvi) intelligenciájára. 2a) és 2b) alapján az olvasó szó szerint számtalan hasonló kifejezést tud alkotni mindkét nyelven. A számítógépes rendszer azonban erre pusztán a felszíni alakok alapján képtelen.

### 2.3 A többszavas kifejezések szerkezete

Az 1b) alatt található kifejezéseket tehát szó szerint fordítani nem lehet, felsorolni pedig vagy nem érdemes vagy nem is lehet. A feladat tehát annak a mintának a megalkotása, amelynek segítségével tetszőleges számban képezhetők illetve megfeleltethetők az ilyen szerkezetű elemek. Szerencsére a kifejezéseken belüli kombinációs megszorítások jól kezelhetők véges állapotú lokális grammatikákkal. A Maurice Gross [2] nevéhez fűződő lexikalizált véges állapotú nyelvtanok implementálására Max Silberstein hatékony eszközt fejlesztett [5],[6], melynek magyar nyelvű alkalmazásáról lásd [9].

A kifejezések szerkezetét elemezve azt találjuk, hogy egyes elemek konkrét szóalakok (mint pl. *éves*, *nyelví*) mások viszont vagy teljesen nyílt osztályt alkotnak (számkifejezések, mint pl. *huszonhat*, *kilenc és fél*) vagy egy terjedelmes, bár felsorolható listát (nemzetiségnévek). A 2. ábra az 1b) kifejezéseket előállító lokális grammatikát mutatja, amely listás felsorolást tartalmazó beágyazott gráfot (Language.grf) használ. Ezzel ekvivalens megoldás, ha a szótárban szemantikai jegyekkel különböztetjük meg a Language.grf elemeit, és a fő gráfban a szemantikai jegyekre hivatkozva definiáljuk a gráf megfelelő helyén előforduló elemek körét (lásd. 3. ábra)

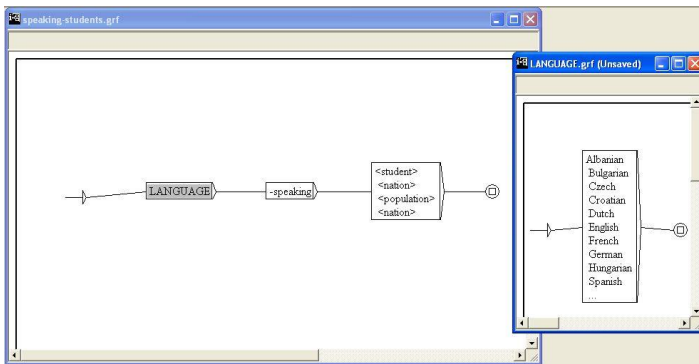
Azt látjuk tehát, hogy a többszavas kifejezések egy része leírható/előállítható egy olyan véges állapotú lokális grammatikával, melynek bizonyos csomópontjait egyedi lexikai elemek töltik ki, a szerkezet más pontjain pedig a lexikai egységek valamilyen osztálya szerepel. Hagyományosan a szófajokat használjuk az egy adott szintaktikai pozícióban behelyettesíthető szavak halmazának képviselőjére. A szintaktikai szerkezeteket, például a főnévi csoportot, a hagyományos grammatikában ismert szófajok kombinációjaként jellemezzük. A feltevés az, hogy miután ezen az általános síkon meghatároztuk a végső nem-terminális elemek kapcsolódásait, a szóosztályokat bármely azonos kategóriájú lexikai elemmel helyettesíthetjük.

Teljesen nyitott szerkezetek leírására ez az eljárás meg is felel. A fenti példák azt bizonyítják, hogy a kifejezések egy részére ez a modell nyilvánvalóan alkalmatlan. A „spanyol nyelvű lakosság” kifejezést nem elégséges „Adj Adj N” kategóriák szekvenciájaként jellemeznünk, hiszen ennek számtalan rosszul képzett, vagy értelmezhetetlen alak is megfelel, mint pl. a 3a) vagy a 3b).

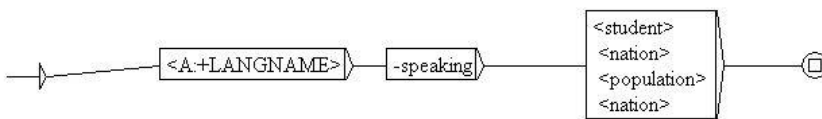
3a) \*nyelvű spanyol lakosság

3b) \*spanyol nyelvű kardántengely

A szintaxis autonómiája nevében természetesen mondhatjuk, hogy a „spanyol nyelvű kardántengely” kifejezés teljesen grammatikus, értelmezést tulajdonítani neki nem is lehetetlen (pl. spanyol nyelvű felirattal ellátott kardántengely?), még ha az nem is lesz egyező a személyek esetében alkalmazott ’spanyol anyanyelvű’ értelemmel. Ugyanakkor azonban a 3a) példa nem szemantikailag, hanem szintaktikailag rosszul formált, pedig ez is „Adj Adj N” szekvencia. Érvelésünk szempontjából közömbös, hogy szintaktikai vagy szemantikai jellegű megszorításról van szó. A lényeg az, hogy a szófajokkal megadott struktúraírás messze nem kielégítő, mert a rendkívül durva kategóriákkal nem tudunk számot adni a szerkezet elemei között fennálló finom megkötésekről. Természetesen a nyelvtechnológiai alkalmazásokban annak a kérdésnek, hogy a kifejezés elemei között fennálló kombinációs megszorítások szintaktikai vagy szemantikai/pragmatikai jellegűek-e semmi jelentősége nincs.



2. ábra Az 1b) kifejezés szerkezetét leíró lokális grammatika



3. ábra Az 1b) kifejezések lokális grammatikája szemantikai jegyek használatával

Az a kérdés sem kell, hogy különösebben foglalkoztasson minket, hogy vajon a szótárba vagy a nyelvtanba tartozó jelenségről van szó. Egyrészt ez részben implementációs kérdés, részben pedig nem kategória hanem fokozatos átmenet (kontinuitás) kérdése. Ugyanazt a technológiát (reguláris grammatikát) alkalmazhatjuk a részben, mint a teljesen nyílt kifejezések leírására. A különbség csupán abban rejlik, hogy a szerkezet elemei csak egyedi lexikai elemekből állnak-e (kötött idiómák) illetve, ha

a lexikai elemek csoportjára utalnak, azt hogyan teszik. Ez utóbbi tekintetben a néhány elemű listától a szófaji kategóriával a lehető legáltalánosabban jellemzett osztályig terjed a skála.

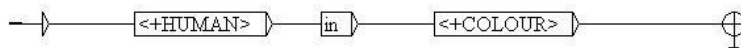
Az itt tárgyalt kifejezésekhez olyan lokális grammatikákat szükséges alkalmazni, amelyek vegyesen tartalmaznak egyedi lexikai elemeket, listákat és szóosztályokat. A véges állapotú gráfok csomópontjain mindig olyan szintű megkötést kell alkalmaznunk, amely a lehető legpontosabban jelöli ki azon lexikai elemek körét, amelyek abban a helyzetben előfordulhatnak. Ha ez egyetlen, adott szóalak, akkor a toldalékolt felszíni alak szerepel, ha a toldalék tetszőleges lehet, de a lexéma nem, akkor a lexéma szintjén specifikáljuk a szerkezeti csomópontot. Másik végletként a szófaji kategóriát alkalmazzuk.

A két véglet (egyedi szóalak, szófaj) között található lexikai csoportok specifikálása különösen érdekes lehet. Akár szemantikai jegyekkel illetve diszjunktív listával, akár véges állapotú automatával határozzuk is meg a minta nyíltvégű elemeit, azok általában intuitíve egy természetes szemantikai osztályt alkotnak. Ilyenek az 1) – 3) példákban is szereplő nemzetnevek és számok. Hasonló szerepet játszhatnak színnevek, testrészek, lásd 4) és 5) példáit.

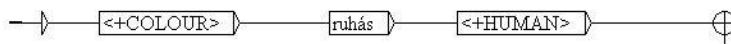
4a) the lady in black  
4b) A fekete ruhás hölgy

5a) baby-/poker-/hard-faced boy  
5b) baba-/póker-/kemény arcú fiú

Például a 4a) és 4b) szerkezetét informálisan meghatározhatjuk a 4. és 5. ábrán látható módon, ami közvetlen implementálható is a Nooj rendszerben (URL: [perso.wanadoo.fr/rosavram](http://perso.wanadoo.fr/rosavram)) Ez a megvalósítás feltételezi, hogy a lexikonban szemantikai jegyeket használunk a színeket valamint a személyeket jelölő szavakra.



**4. ábra** Lokális grammatika a 4a) kifejezésre szemantikai jegyek használatával



**5. ábra** Lokális grammatika a 4b) kifejezésre szemantikai jegyek használatával

A 4. és 5. ábrán látható szemantikai jegyek használata csak az első lépcső a lexikon kiépítésében. Egy magasabb szervezettségű lexikonban a szavak megfelelő csoportját nem egyedi szemantikai jegyek vagy azok halmaza jelöli ki, hanem a jegyek típushierarchiába szervezett szerkezete, amely öröklődést is lehetővé tesz.

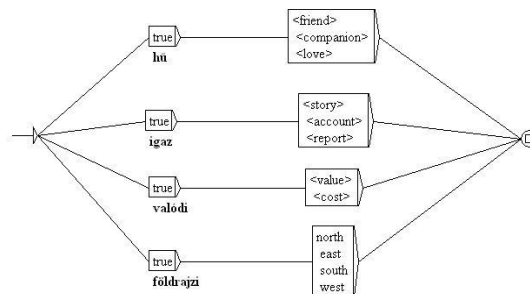
Az eddigi példák szándékosan egyszerűek voltak, de az az állításunk, hogy az általuk illusztrált korlátozottan nyitott szerkezetű kifejezések korántsem marginális szerepűek a nyelv lexikai szerkezetében és szerkezetük sem olyan triviális, mint az eddig említett példák. Tekintsük például a 7. ábrán szereplő lokális grammatikát, amely az időkifejezések szerkezetének legfelső szintjét mutatja be. A 8. ábra az „oraperc” nevű algráf egy részét tartalmazza<sup>69</sup>.

### 3. A lokális grammatikák alkalmazásai

Végezetül a fent bemutatott, korlátozottan produktív kifejezések lokális grammatikáinak három nyelvtechnológiai alkalmazását mutatjuk be.

#### 3.1 Szemantikai egyértelműsítés

A lokális grammatika kiváló eszközzel szolgálhat egy-egy kifejezésben szereplő lexikai egység egyértelműsítésére, amit a 6. ábrán szereplő transzducser példáz. Az angol „true” szónak az adott kontextusban érvényes magyar megfelelőit a gráf kimenete szolgáltatja, melyet a csomópont alatt találunk.



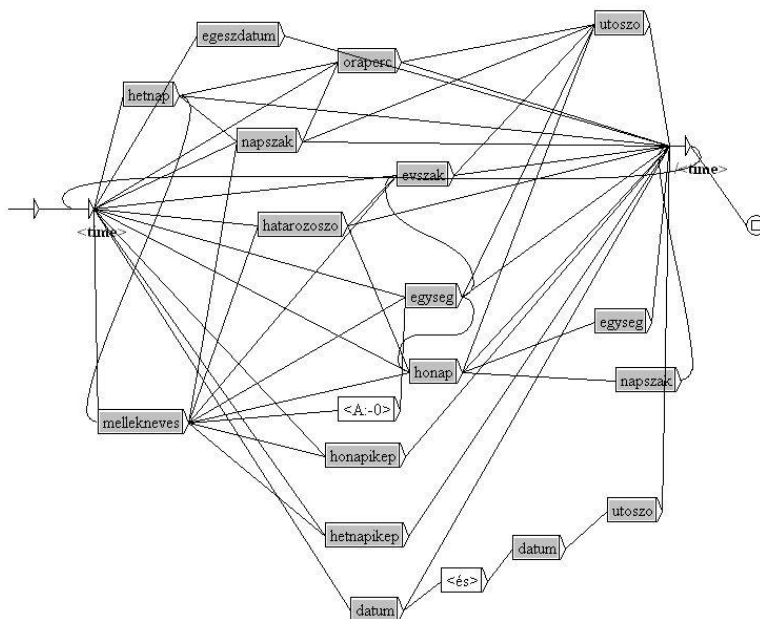
6. ábra Szemantikai egyértelműsítés lokális grammatikával

<sup>69</sup> A gráfokat Gábor Kata készítette munkatársaival az MTA Nyelvtudományi Intézet Korpusz-nyelvészeti Osztályán.



### 3.2 Részleges felszíni gépi fordítás

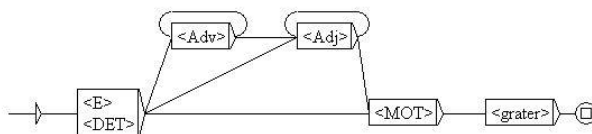
A felszíni elemzés (shallow parsing) mintájára szerkeszthetünk olyan transzdúcert is, amely a felszíni forrásnyelvi minta kimeneteként annak célnyelvi megfelelőjét adja. Erre látunk példát a 9. ábrán, ahol egy olyan grammatika látható, melynek segítségével a „trade in dollar” kifejezéseket a megfelelő „dollárkereskedelem” kifejezéssel, a „trade in London” stb. kifejezéseket pedig a „londoni kereskedés” megfelelővel fordíthatjuk le.



7. ábra Időkifejezések lokális grammatikája



az üres elemet jelöli. Az ábrán a <DET> szimbólummal diszjunktív kapcsolatban szerepel, ami opcionálissá teszi a csomópontot a grammatikában.



**10. ábra** Automatikus lexikai elsajátítás lokális grammatikával

## 4. Összefoglalás

A jelen dolgozatban a többtagú kifejezések egy kevésbé vizsgált fajtájára hívtuk fel a figyelmet, amelynek meghatározó jellemzője, hogy részlegesen produktívak, azaz szerkezetük elemei között szerepelnek olyanok, amelyek egy többé-kevésbé nyitott szóosztállyal jellemezhetők. A számítógépes alkalmazások, mint például a gépi fordítás lexikonja számára az ilyen kifejezések explicit jellemzése szükséges, nem elégedhetünk meg a hagyományos lexikográfia utalásos, példalódzós módszerével. Ez teszi szükségessé az ilyen részleges nyitott kifejezések lokális grammatikák segítségével történő kezelését, melyre több példát említettünk.

A kifejezések elemei között olyan finom szintaktikai, de túlnyomórészt szemantikai kombinációs megszorítások vannak, amelyekre a szófaji kategóriák túl durvának bizonyulnak. A természetes szemantikai osztályt alkotó szavak (mint pl. színnév, nemzetnév, testrészt, szám stb.) csoportját a lexikonban definiált hierarchikusan szervezett, öröklődő jegyek segítségével lehet legjobban megragadni.

A dolgozat végén bemutattunk egy komplex példát az időkifejezések grammatikájából valamint a lokális grammatikák alkalmazását szemantikai egyértelműsítésre, mintaillesztéses felszíni fordításra valamint automatikus lexikonfejlesztésre.

## Bibliográfia

1. Bolinger, D. (1965). "The Atomization of Meaning." *Language* **41**: 555-573.
2. Gross, M. (1997). The Construction of Local Grammars. in Y. S. Emmanuel Roche (szerk.) *Finite State Language Processing*. MIT Press: 329-352.
3. Oravecz, Cs. et al. 2004 Többszavas kifejezések számítógépes kezelése in Alexin, Z., Csendes, D. (szerk.) *II. Magyar Számítógépes Konferencia*. Szeged. *SZTE Informatikai Tanszékcsoport*: 141-150

4. Sag, I. et al. 2002 Multiword Expressions: A Pain in the Neck for NLP. in Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics CICLING 2002: 1--15,
5. Silberztein, M. (1993). Dictionnaires électroniques et analyse automatique de textes: le systeme INTEX. Paris, Masson.
6. Silberztein, M. (1999). "Text Indexation with INTEX." Computers and the Humanities **33**(3): 265-280.
7. Sinclair, J., (szerk.) (2004). Collins Cobuild Advanced Advanced Learner's English Dictionary. Glasgow. HarperCollins Publishers
8. Summers, D. (szerk.) (2003). Longman Dictionary of Contemporary English. Harlow. Pearson Education Ltd.
9. Várad, T., Gábor K., (2004) A magyar INTEX fejlesztéséről in Alexin, Z., Csenedes, D. (szerk.) II. Magyar Számítógépes Konferencia. Szeged. SZTE Informatikai Tanszékcsoport: 3-10
10. Wehmeier, S., (szerk.) (2005). Oxford Advanced Learner's Dictionary. Oxford, Oxford University Press.