

Angol időhatározói NP-k a gépi fordításban

Ugray Gábor, Ujvárosi Gábor

MorphoLogic Kft.
{ugray,ujvarosi}@morphologic.hu

Absztrakt. Az időhatározói NP-k felismerése az elemzés során kulcsfontosságú a helyes gépi fordítás szempontjából, s mivel sajátos belső szerkezettel illetve disztribúcióval rendelkeznek, a szerzők ezek lefedésére különálló, kézzel kódolt szabályokat hoztak létre. Időhatározó-nyelvtanunk jelenleg több mint 800, példamondatokkal annotált szabályt tartalmaz. A témában rendelkezésre álló irodalom igen szűkös, különösen ami a csupasz időhatározói NP-ket illeti. A munka során szerzett tapasztalataink hatására számos általánosan határozószóként számon tartott szó besorolását megváltoztattuk, és az NP-knek és PP-knek az angol nyelvtanokban hagyományosan éles elkülönítése is megkérdőjeleződött.

1. Bevezető

A mód-, hely illetve időhatározói értelmű prepozíciós csoportok (PP-k) megkülönböztetése a szintaxis számos alkalmazásánál nem bír különleges jelentőséggel, így nem meglepő, hogy az angol nyelv egyszerű generatív nyelvészeti leírásai (Radford 1988) az alábbiakhoz hasonló szabályokból építkeznek: $V' \rightarrow V (NP) (PP)$ és $V'' \rightarrow V' (PP)$. A gépi fordítás szempontjából azonban kulcsfontosságú, hogy a szabad módosítóként álló PP-t (pl. *in June*) helyesen értelmezzük időhatározóként a – legalábbis a mi rendszerünkben – elsődlegesebb helyhatározó helyett.

A szabad módosítóként álló PP főnévi fejének lexikai tulajdonságai nyilvánvaló módon kitétetett szerepet játszanak abból a szempontból, hogy a PP időhatározói funkciót tölt-e be a mondatban. Az *at the bus stop* szerkezet a főnévi fej lexikai tulajdonságaiból kifolyólag lesz helyhatározó, míg az *at six o'clock* a *six o'clock* NP tulajdonságai miatt működik időhatározóként.

Az alábbi időhatározói szerkezeteket megvizsgálva kitűnik, hogy a bennük szereplő NP-k igen különleges tulajdonságokkal rendelkeznek:

1. after July 4
2. at 6.30 am
3. on May 27, 1978
4. in 2003
5. before last year

Világos, hogy az 1-4. példában szereplő, dátumot ill. időt leíró ún. *clock-calendar* NP-k sajátos szerkezettel bírnak, amit a „produktív”, a vonásszámot bal- illetve jobb-

oldali komplementumok és módosítók hozzávételével növelő szabályok nem képesek leírni. Ebből az alapvető megfigyelésből kiindulva a produktív NP-nyelvtan mellett elkezdtünk kézzel írt szabályokban olyan NP-eket kódolni, amelyek nagyon kevés általános, aluspecifikált elemet tartalmaznak és „kilapított” struktúrákat írnak le, azaz jobboldalukon többnyire lexikálisan megkötött terminálisokat tartalmaznak. A fenti 3. példát leíró szabály, vázlatosan:

(Sz1) NP = N(lex="May") + NUM + PUNCT(lex="comma") + NUM

Továbbmenve, az 5. példában szereplő NP-nek az angolban egyáltalán nem „szabadna” előfordulnia, hiszen megszámlálható főnév egyes számban nem állhat determináns nélkül.¹

2. Csupasz adverbialis NP-k²

Az időhatározói PP-kben előforduló NP-k szokatlan szerkezete valójában csak a nehezségek kezdete. Mint azt korábban már megfigyelték (Larson 1985; McCawley 1988; Bresnan 2000), bizonyos főnévi konstrukciók prepozíció nélkül is betölthetnek határozói szerepet:

6. A strange man knocked at our door one day.
7. The country's GDP grew by 4.5 percent last year.
8. John was reading the papers all morning.
9. The president said Sunday he didn't support these plans.

Larson és McCawley saját nyelvészeti paradigmájukon belül megpróbálják leírni ezt a jelenséget. Az egyik megoldás, hogy a fejként álló főnevet olyan lexikai jellemzővel ruházzák fel, ami révén az egész NP automatikusan esetet kap; a másik, hogy ezeket a csoportokat ϵ -prepozíciós PP-nek tekintik. Mi egyik értelmezés mellett sem foglalunk állást. Célunk, hogy saját rendszerünkön belül a lehető legtöbb időhatározót³ lefedjük, ezért a kézzel írt szabályok közé elkezdtünk csupasz NP-adverbiálisokat is felvenni.

¹ Említésre méltó, hogy a *next* ill. *last* melléknevek egyfajta korlátozott determinánsi funkcióval is rendelkeznek. Míg a *next door* és a *last year* teljesen helyesek, a **small door* és a **long year* már nem grammatikus NP-k. A helyzetet tovább bonyolítja, hogy csak bizonyos főneveteket determinálhat ez a két melléknév, ld. **next house*, **last moment*.

² A *csupasz adverbialis* NP kifejezést az angol nyelvű szakirodalomban alkalmazott *bare adverbial NP* megfelelőjeként használjuk.

³ A *határozó* kifejezést az angol *adverbial*-hez hasonlóan tág értelmezésben használjuk, azaz így nevezünk minden határozói értelmű szerkezetet, nem csak a határozószókat ill. azok kiterjesztéseit.

3. Kategorizálás

Az olyan kifejezések, mint a *one day*, *last year* és *all morning* nyilvánvalóan a főnévi csoportokra jellemző szerkezettel rendelkeznek, így természetes döntésnek tűnt, hogy akként is írjuk le őket, bár ez felveti azt a problémát, hogy hogyan tegyük lehetővé egy NP-nek, hogy időhatározói módosítóként viselkedhessen. Az ilyen NP-k disztribúciója további érveket szolgáltatott döntésünk helyessége mellett:

10. I didn't want to tell you the secret before tomorrow.
11. I lived here during last year.
12. This morning was fine.

A 10. ill. 11. mondatban a *tomorrow* és a *last year* PP-kben szerepelnek, az 12. példában pedig a *this morning* a mondat névszói alanya. Ezt a viselkedést igen nehéz lenne megmagyarázni, ha akár határozói, akár prepozíciós csoportként kategorizálnánk őket, ami a másik két lehetőség.

Még egy lépéssel továbbvive ezt a gondolatmenetet, az olyan szavakat, mint a *yesterday*, *today* és *tomorrow* szintén főnévként – illetve a jelenlegi leírásban NP-ként – kell kezelnünk.⁴ Hagyományosan ezzel szemben feltehetőleg habozás nélkül határozószóként tüntetnénk fel őket a lexikonban.

4. A nyelvtan

Időhatározó-nyelvtanunk jelenlegi állapotában több mint 800 szabályt tartalmaz. Tesztelési célból minden szabályhoz legalább egy példamondat tartozik. Mint fent említettük, az esetek nagy részében „lapos” NP-eket írunk le,⁵ azaz a csoport belső szerkezetéről nem, vagy csak minimális mértékben veszünk tudomást. A produktív szabályokkal való átfedésnek köszönhetően sok ilyen NP-nek két elemzése is születik, azonban a határozó-nyelvtan által létrehozott elemzéseket egy kitüntetett jeggyel megkülönböztetjük, s megfelelő felülbírással a teljes mondatnak nem keletkezik két értelmezése.

Az időhatározói NP-k csökevényes struktúrájú, kevés általánosítást tartalmazó, lexikálisan specifikált szabályokkal történő kódolásának fő oka, hogy fordításuk nagy mértékben megjósolhatatlan. Ennek tükrében az alkalmazott „nyers erő módszere” tűnt a leggyümölcsözőbb útnak. Corley és Haywood (2000) ezen kívül rámutatnak,

⁴ Érdekes megfigyelni, hogy bizonyos csupasz adverbialis NP-k szerepelhetnek mondat névszói alanyként ill. prepozíciós csoportban (*Yesterday was fine*, *She lived elsewhere before last year*), más, nyilvánvalóan NP-szerkezetet mutató kifejezések ezt nem engedik meg (**All morning was fine*).

⁵ Ez természetesen nem vonatkozik azokra az esetekre, ahol a kifejezés disztribúciója nem mutat főnévi csoportra jellemző jegyeket, mint pl. az idiomatikus *every now and then*. Az ilyeneket határozói csoportként kódoltuk.

hogy a csupasz adverbialis NP-k különleges helyet helyet élveznek az emberi nyelvtani reprezentációban is.

A nyelvtanunk egyéb területein is alkalmazott elv, hogy sok lexikális információt a terminálisok helyett magasabb szinteken kódolunk. Egy adott ige szubkategorizációit például VP-t létrehozó szabályok formájában soroljuk fel, amik az igére lexikai kikötést tartalmaznak, és jobboldalukon szerepelnek annak komplementumai. Az elemzőszabályhoz tartozó generálósabályokban a komplementumok megkapják a megfelelő magyar esetet, névutót stb. Mindezt az információt jóval egyszerűbb különálló, a komplementumokat is magukban foglaló, lexikálisan specifikált VP-s szabályok formájában leírni, mint az igéhez tartozó egyetlen lexikon-bejegyzésbe sűríteni. Ugyanezt az elvet követjük, amikor egyedi, bonyolult lexikai jegyeket hordozó főnevek helyett egész NP-t tartalmazó szabályokat írunk le.

Az alábbiakban néhány vázlatos szabályon keresztül bemutatjuk az időhatározó-nyelvtanban kódolt NP-k fontosabb attribútumait. A példák csak az elemző oldali részt tartalmazzák, a generáló szabályokra helyszűke miatt nem térünk ki.⁶

(Sz2) *NP=last+year:0208220945-1
EN.NP[temp=YES, mayadv=YES, timep=YES, genadj=YES] =
ADJ(lex="last") + N(lex="year", num=SG)

A (Sz2) a *last year*-t reprezentálja. A *temp* a megkülönböztető jegy, aminek YES értéke azt jelenti, hogy az adott NP-t az időhatározó-nyelvtanhoz tartozó szabály hozta létre. A *mayadv* azt fejezi ki, hogy csupasz adverbialis NP-ről van szó. Az ilyenek megengedik a következő szabály működését, ami már egy időhatározói szerepű szimbólumot hoz létre:

(Sz3) *ADVP=NP:0206261835-1
EN.ADVP[lexical=YES, ppreo=YES, pfinal=YES, prnp=NO, pinit=YES,
type=TEMP, ttemp=YES] = NP(temp=YES, mayadv=YES)

A *genadj* szintén az NP disztribúciójáról tartalmaz információt. Nyelvtanunkban az időhatározói NP-k illetve a belőlük képződött PP-k négy helyen fordulhatnak elő: i) VP jobboldali módosítójaként, ld. (Sz3) és 13.; ii) NP jobboldali módosítója, ld. (Sz8) és 14.; iii) NP baloldali módosítója, ld. (Sz6) és 15.; valamint iv) NP birtokos determinánsa, ls. (Sz4) és 16. A *genadj* YES értékének jelentése, hogy az NP állhat állhat birtokos determinánsként.

13. He arrived last year.
14. The excursion last year was a lot of fun.
15. The six o'clock train was late again.
16. Last year's excursion was a lot of fun.

⁶ A szabályok első sora fejléc, a nyelvtani értelemben vett információt a második sor tartalmazza, amely a nyomdai korlátok miatt itt több sorba törik. (Sz2) a megszokott írásmóddal az NP → ADJ N alakú szabálynak felel meg. A szögletes zárójelek közötti értékadásokban a létrejövő szimbólum jegyeinek alapértelmezés szerinti értékeit bíráljuk felül; a jobboldali szimbólumok után álló kerek zárójelek között pedig azok jegyeire teszünk kikötéseket.

Az egyik szabály, ami lehetővé teszi, hogy időhatározói NP-k más NP-k birtokos determinánsaként szerepelhessenek az (Sz4):

(Sz4) *NPX=NP(adj)+S+NM:0304170339-10
EN.NPX[...] = NP(genadj=YES, temp=YES) + GENS + NM(...)

Fontos megkülönböztetni ezt az esetet a *the man's wife*-hoz hasonló „valódi” birtokos szerkezetektől. Míg az utóbbiak magyarban is birtokos szerkezetre fordulnak, az előbbiekből jelzői szerkezetet képzünk (*a tavalyi kirándulás*).

A mayadj jegy azt fejezi ki, hogy az NP egy főnév baloldali módosítójaként állhat, mint a 15. példában. Az itt szerepet játszó szabályok:

(Sz5) *NP=NUMX+oclock:0206081108-11
EN.NP[temp=YES, mayadj=YES, timep=YES, timet=HR, ofradj=YES, timeprep="at"] = NUMX(numtype=CARD) + N(lex="o'clock")

(Sz6) *ADJY=NP:0208081727-1
EN.ADJY[type=TIME] = NP(temp=YES, mayadj=YES)

(Sz7) *NN=ADJP+NX:0205291926-1
EN.NN[...] = ADJP + NX(...)

A csupasz adverbialis NP-k jobboldali módosítóként való elemzését lehetővé tevő szabály:

(Sz8) *RADJ=timeNP:0302110214-2
EN.RADJ[radjuid="adv", hupos=LEFT] = NP(temp=YES, mayadv=YES)

Azok az időhatározói NP-k, amelyek prepozícióval együtt fordulnak elő, többnyire három csoport valamelyikébe esnek aszerint, hogy a „mikor?” kérdésre adott válaszhoz az *in*, *or* ill. *at* közül melyiket választják. Ennek illusztrálására szolgáljon az alábbi három példa, valamint a hozzájuk tartozó szabályok:

17. The accident happened on August 23.

18. There was no rain in the spring.

19. John's train arrived at 6 am.

(Sz9) *NP=MONTH+DAY:0207111442-2
EN.NP[temp=YES, mayadv=YES, mayadj=YES, timet=DATE, timep=YES, timeprep="on", num=SG, pers=P3, ofradj=YES] = NX(month=YES, num=SG) + NUMX(numtype=CARD, middle=YES)

(Sz10) *NP=the+season2:0208281300-1
EN.NP[temp=YES, timep=YES, timeprep="in"] = DET(detttype=DEF) + NX(season=YES, lex!="winter", lex!="summer")

(Sz11) *NP=NUMX+am:0206061754-2
EN.NP[temp=YES, mayadj=YES, timet=HR, timep=YES, timeprep="at"] = NUMX + ADV(lex="am")

(Sz12) *ADVP=PREP+NP(timep):0207021632-1
EN.ADVP[lexical=YES, pfinal=YES, pinit=YES, prnp=NO, type=TEMP, ttemp=YES] (PREP.lex=NP.timeprep) = PREP + NP(temp=YES)

A timeprep jegyben tároljuk az NP által választott prepozíciót. Amennyiben üresen marad – s ez az alapértelmezett érték –, a (Sz12) nem sülhet el, mert a baloldalra írt feltétel, hogy a prepozíció szótári alakja egyezzen meg az NP timeprep jegyében tárolt értékkel, nem teljesül. Létezik egy hasonló szabály, ami révén a prepozíciót igénylő NP-k előfordulhatnak főnév jobboldali módosítójaként (ld. 20. példa), bár tapasztalatunk szerint nem minden VP-módosítójaként megfigyelhető PP működhet főnév jobboldali módosítójaként is.

20. The meeting at 6 am was quite a thrill.

Bizonyos prepozíciók (például a *before*, *for*, *during* stb.) előfordulása az NP más lexikai tulajdonságától függ.

(Sz13) *ADVP=after:0206260909-2
EN.ADVP[lexical=YES, pfinal=YES, pinit=YES, prnp=NO, type=TEMP, ttemp=YES] = PREP(lex="after") + NP(temp=YES, timep=YES)

(Sz14) *ADVP=for+NX(dura):0303071309
EN.ADVP[lexical=YES, prnp=NO, pinit=YES, pfinal=YES, ppreo=YES, type=TEMP, ttemp=YES] = PREP(lex="for") + NP(temp=YES, dura=YES)

A timep jegy durván azt fejezi ki, hogy a szóban forgó NP egy „időpontra” vonatkozik, míg a dura jelentése, hogy „időtartamot” ír le.

A *for* és *in* prepozíciók különleges figyelmet igényelnek, mivel kétféle szemantikai értelmezésük is létezik. (Hitzeman 1996) amellet érvel, hogy ezek függetlenek az igeidőtől, a mi céljainkra azonban „elég jó” heurisztikának megfelel, hogy megkülönböztetésül az igeidőre támaszkodjunk. Az elemzési oldalon pedig a két értelmezés között szintaktikai szempontból egyáltalán nem mutatkozik különbség.

5. Hivatkozások

1. Bresnan, J.: *Lexical-Functional Syntax*. Blackwell Publishers, Oxford (2000)
2. Chomsky, N.: *Remarks on Nominalization*. In: R. Jacobs and P. Rosenbaum (szerk.): *Readings in English Transformational Grammar*. Blaisdell Publishing, Waltham, Massachusetts (1970)
3. Corley, M. & Haywood, S.: *Parsing modifiers: The case of bare-NP adverbs*. In: *Proceedings of the 21st annual meeting of the Cognitive Science Society*, 126-131. Vancouver (2000)
4. Hitzeman, J.: *Semantic partition and the ambiguity of sentences containing temporal adverbials*. Online: <http://www.hcrc.ed.ac.uk/publications/rp-77.ps.gz> (1996)
5. Komlósy, A.: *A lexikai-funkcionális grammatika mondattanának alapfogalmai: Segédkönyvek a nyelvészet tanulmányozásához VII.*; Kiss Gábor (szerk.) *Nem transzformációs nyelvtanok I.* Tinta Könyvkiadó, Budapest (2001)
6. Larson, R.: *Bare-NP Adverbs*. In: *Linguistic Inquiry* 16, 595-621. (1985)
7. McCawley, J D.: *Adverbial NP's: bare or clad in see-through garb?* In: *Language* 64, 583-90. (1988)
8. Radford, A.: *Transformational Syntax*. Cambridge University Press, Cambridge (1988)