

# English Adverbial NPs of Time in Machine Translation

Gábor Ugray and Gábor Ujvárosi  
MorphoLogic Ltd.  
8 Késmárki u., 1118 Budapest, Hungary  
{ugray,ujvarosi}@morphologic.hu

## Abstract

The authors are participants of a project aimed at developing an English-Hungarian machine translation (MT) system. Recognizing time adverbial NP's in the parse side is crucial for correct MT, and the fact that they exhibit peculiar properties in their structure and distribution has prompted the authors start hand-coding a separate set of NP rules for them. At the time of writing they have over 700 rules annotated with test examples. Literature on the issue is scarce especially with regard to bare adverbial NP's, and their experiences have lead them to conclusions about the PoS of some words traditionally treated as adverbs, and the role of NP's and PP's in their English syntax.

## 1 Introduction

The distinction between PP adverbials of manner, space and time may not be of great importance for some applications of syntax, thus it is not surprising that rudimentary generative grammars of English (Radford 1988) operate with rules such as  $V' \rightarrow V$  (NP) (PP) and  $V'' \rightarrow V'$  (PP). For the purposes of machine translation, however, correctly recognizing an adjunct PP such as *in June* as an adverbial of time instead of the most common interpretation – in our case – as a location adverbial is a crucial issue.

It seems an obvious observation to us that it is the lexical properties of the adjunct PP's noun head that play the chief role in determining whether the PP will function as an adverbial of time. *At the bus stop* is interpreted as an adverbial of space because of the lexical properties of its noun head, *bus stop*, while *at six o'clock* is an adverbial of time because of the properties of the NP *six o'clock*.

Upon examining the following set of temporal adverbials it turns out that the NP's in them have some very peculiar characteristics:

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

Clearly, clock-calendar NP's such as (1)-(4) have a very special structure that cannot be accounted for using our “productive” NP-rules that increase the bar number of N by subsequently including left and right complements and adjuncts. This basic observation has lead us to start hand-coding a parallel set of NP rules that contain very little generalization and create “flattened” structures by effectively listing lexically specified terminal symbols in their right-hand side. A skeletal example of such a rule<sup>1</sup>, describing (3) above, is:

(R1) NP = N(lex=“May”) + NUM +  
PUNCT(lex=“comma”) + NUM

The NP in (5), furthermore, should be perfectly illegitimate in English, considering the constraint that no count noun may stand without a determiner in singular.<sup>2</sup>

## 2 Bare adverbial NP's

The observation that the NP's in temporal PP adverbials exhibit structural irregularities is barely scratching the surface of the matter. As has been noted in the literature (Larson 1985; McCawley 1988; Bresnan 2000), some nominal constructions can function as adverbials even in the absence of a preposition:

---

<sup>1</sup> The comma-separated expressions within parenthesis in the right-hand side of the rule mean a conjunction of conditions for the features of the symbols

<sup>2</sup> It is worth noting that “next” and “last” as adjectives appear to play a special role as determiners in English. While *next door* and *last year* are perfectly good phrases, *\*small door* and *\*long year* are ungrammatical without a determiner. To complicate matters further, certain nouns appear to be able to participate in these constructions while others do not: *\*next house*, *\*last moment*.

- (6) A strange man knocked at out 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 and McCawley both attempt to account for this phenomenon within their own linguistic paradigm. One solution is to include a lexical feature for the head nouns in question that automatically assigns oblique case to the NP in which they participate; the other is to treat these structures as PP's with an  $\epsilon$  preposition. We do not intend to argue either against or for any of these assumptions. Our aim is to correctly recognize as many of the temporal adverbials within our framework as possible, therefore we set out to include bare-NP adverbials in our hand-coded set of rules.

### 3 Categorization

Expressions such as *one day*, *last year* and *all morning* clearly exhibit the structure of a noun phrase, therefore it was a natural decision to encode them as such, even though this implies the problem of how to allow an NP to function as an adverbial adjunct in our grammar. Further evidence for treating such constructions as NP's comes from their distribution:

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

In (10) and (11), *tomorrow* and *last year* participate in PP's, and in (12) *this morning* acts as the NP subject of the sentence. Clearly this would be very difficult to explain if we categorized them as either adverbial or prepositional phrases, which are the two other candidates.

Taking this line of thought one step further, words such as *yesterday*, *today* and *tomorrow* should clearly be treated as nouns, or – in our interpretation – as noun phrases proper. Traditional analyses would no doubt intuitively list them as adverbs<sup>3</sup> in the lexicon. To account for *before tomorrow*, then, there would remain only two very cumbersome solutions. The first is to allow a rule like

<sup>3</sup> We adopt here the convention of using the term *adverb* as a particular PoS, as in the case of *quickly*, for instance, and applying *adverbial* in a broader sense to cover *here*, *on the roof* and *last night* as well.

- (R2) PP = PREP + ADV

The second is to alternatively list them as nouns as well in the lexicon, but that has the counterintuitive consequence that *tomorrow* in *I didn't want to tell you the secret before tomorrow* is interpreted as a noun, while the same word in *He will arrive tomorrow* is interpreted as an adverb. To us it seemed a much more natural solution to add a feature to some NP's that allows them to function as adverbials.

### 4 The grammar

At this point it is necessary to say a few words about the concepts used in our grammar to be able to explain how our time adverbial NP's function. The grammar is essentially a phrase structure grammar, and in the analysis of noun and verb phrases we loosely follow the principles of X-bar theory. The first extension of N, which we term NX, includes the left and right complements of the noun. A further extension, NN, includes left and right adjuncts. We have created a separate level, NM, to include right clausal adjuncts. The level NPX corresponds to the largest extension (including determiner). NP's, then, are reserved for coordinated NPX's.

As for the verb phrase, we label the content verb and its subcategorization frame VP, while a VP plus its adjuncts is termed PRED. In most cases, PRED includes all functional verbs (have, be etc.) except the ones with the tense marker. The two notable exceptions are the simple present and simple past tenses, where the content verb proper bears the tense marker. PRED also includes adverbials whose position is defined as pre-verbal or between the auxiliary and the verb.

We wrap up NP's and PP's in what could be termed "functional" categories, so instead of writing rules as (R3) and (R4) below, we employ ones principally similar to (R5) through (R7), although of course in reality they are much more restricted:

- (R3) PRED = VP + PP
- (R4) PRED = VP + ADV
- (R5) PRED = VP + ADVP
- (R6) ADVP = PP
- (R7) ADVP = ADV

For right adjuncts in our "productive" (unrestricted) NP grammar, we use the category RADJ, which is often an envelope of ADVP itself, based on distributional evidence such as (13) and (14) below:

- (13) The man was walking *in the garden*.  
 (14) The man *in the garden* was walking.

## 5 The temporal adverbial NP grammar

At the time of writing, we have over 700 hand-coded rules that represent time adverbial NP's and their wrap-up in functional categories. All of the rules are annotated with several example sentences that serve to test the rules' functionality. As mentioned above, we took the decision to flatly encode whole NP's and ignore their internal structure. The NP's recognized by these rules and by the productive ones just described often overlap, but time adverbial NP's bear a special feature that distinguishes them from their productively created counterparts.

Our chief motivation for this decision was the large unpredictability in the structure of these phrases. In the light of this variability our brute force method seemed the most fruitful approach. Further evidence is found in (Corley-Haywood 2000) that bare NP adverbials occupy a special place in the representation of the human parser as well.

It is also a general principle in our grammar that we encode a lot of lexical information on higher levels. Bear in mind that our goal is machine translation, and the translations of similar structures in the source language can be highly unpredictable. It is, for instance, much more convenient to list all possible subcategorizations of a given verb in the form of several rules that have a VP in their left-hand side, specify the given verb and its complements in the right-hand side, and define the adequate case assignment to the NP's in the Hungarian translation<sup>4</sup> than to encode all of this information in a single lexical entry for the verb. The fact that we encode whole adverbial NP's instead of specifying a lot of convoluted features for individual nouns follows from the same principle.

We will now demonstrate the features with which we furnish our NP's through a few representative rules<sup>5</sup>.

<sup>4</sup> Instead of prepositions, which are completely missing from the language, Hungarian heavily relies on a complex case system and a set of postpositions to express subcategorization information on the nominal complements of verbs.

<sup>5</sup> As the chief aim of this paper is to describe the representation of time adverbial NP's in the source language, English, we omit the parts of rules that define the translation of these phrases.

- (R8) \*NP=last+day:0208220945-1  
 EN.NP[temp=YES, mayadv=YES,  
 timep=YES, genadj=YES] =  
 ADJ(lex="last") + N(lex="day",  
 num=SG)

(R8) represents *last day*. *Temp=YES* is the distinguishing feature which expresses that this NP was created by the set of rules responsible for time adverbial NP's. *Mayadv=YES* specifies that the NP in question can function as a bare adverbial NP by triggering the following rule:

- (R9) \*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)

*Genadj=YES* also contains information about the distribution of the NP. In our grammar, a time adverbial NP (or PP) can fulfill one or more of the following four roles: i) adjunct of a VP, as in (R9) and (15); ii) right adjunct of an NP, as in (R14) and (16); iii) left adjunct of an NP, as in (R12) and (17); and iv) possessive determiner of an NP, as in (R10) and (18). *Genadj=YES* is an indicator that the NP can function as iv).

- (15) He arrived *last day*.  
 (16) The party *last day* was a lot of fun.  
 (17) The *six o'clock* train was late again.  
 (18) *Last day's* party was a lot of fun.

One of the rules that allow time NP's to function as possessive determiners of another NP is (R10):

- (R10) \*NPX=NP(adj)+S+NM:0304170339-10  
 EN.NPX[...] = NP(genadj=YES,  
 temp=YES) + GENS + NM(...)

It is important to distinguish these constructions from possessive structures such as *the man's wife*. Whereas the latter is translated as a genuine possessive, the NP in *last day's party* becomes an adjectival phrase in the translation:

- (20) a            tegnapi                            buli  
       a            tegnapi            -i            buli  
       the[DET] yesterday[N] [N->ADJ] party[N]

The feature *mayadj* expresses that the NP can play the role of a left adjunct in a noun phrase, as in (17). The rules involved are:

- (R11) \*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")
- (R12) \*ADJY=NP:0208081727-1  
 EN.ADJY[type=TIME] = NP(temp=YES,  
 mayadj=YES)
- (R13) \*NN=ADJP+NX:0205291926-1  
 EN.NN[...] = ADJP + NX(...)

As for bare NP adverbials serving as right adjuncts in an NP, we have (R14) below:

- (R14) \*RADJ=timeNP:0302110214-2  
 EN.RADJ[radjud="adv",  
 hupos=LEFT] = NP(temp=YES,  
 mayadv=YES)

Time adverbial NP's that only occur with a preposition tend to fall into one of three groups based on the preposition they select to form a PP adverbial that answers the question "when." This can be illustrated with the following sentences and corresponding rules:

- (21) The accident happened on *August 23*.  
 (22) There was no rain in *the spring*.  
 (23) John's train arrived at *6 am*.
- (R15) \*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)
- (R16) \*NP=the+season2:0208281300-1  
 EN.NP[temp=YES, timep=YES,  
 timeprep="in"] = DET(dettype=DEF)  
 + NX(season=YES,  
 lex!="winter", lex!="summer")
- (R17) \*NP=NUMX+am:0206061754-2  
 EN.NP[temp=YES,  
 mayadj=YES, timet=HR, timep=YES,  
 timeprep="at"] = NUMX +  
 ADV(lex="am")
- (R18) \*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)

The feature *timeprep* defines the preposition of choice of the NP to form a PP adverbial answering the basic question "when." If it is left blank, as in the case of bare NP adverbials, rule (R18) will not fire. A similar rule exists to account for PP's in the right adjunct

position in NP's as in (24), although it has been our experience that not all of the PP's that can function as verbal adjuncts will be nominal adjuncts as well.

- (24) The meeting at *6 am* was quite a thrill.

There is a number of prepositions such as *before*, *for*, *during* etc. whose occurrence depends on different lexical properties of the NP.

- (25) I will finish this task after *July 4, 2002*.  
 (23) The Romans killed barbarians for *centuries*.

- (R19) \*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)

- (R20) \*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)

The feature *timep* expresses that the NP in question refers to "a point in time," while *dura* shows that it represents an interval.

The prepositions *for* and *in* require special attention because their semantic interpretation depends on the tense of the verb which they modify.<sup>6</sup>

- (24) I saw her *for a month*.  
*Egy hónapig* láttam őt.  
 (25) I have been seeing her *for a month*.  
*Egy hónap óta* láttam őt.

For the syntactic analysis of English sentences, however, this distinction doesn't appear to be relevant in our system, so we do not have any provisions to distinguish between the two interpretations in the source side of our rules.

Finally, there are many expressions which might be argued to exhibit the structure of an NP or PP, but whose occurrence is very limited. They are often also idiomatic, so we have decided to encode them directly as ADVP instead of NP's. An example would be *every now and then*, which is encoded in the following rule:

<sup>6</sup> (Hitzeman 1996) argues that the different interpretations are in fact independent from the tense of the sentence, but for our purposes basing out judgement on it appears to yield satisfactory results.

(R21) \*ADVP=every+now+and+then:0207031  
 EN.ADVP[lexical=YES, prnp=NO,  
 pinit=YES, pfinal=YES, ppreo=YES,  
 type=FREQ] = DET(dettype=EVERY) +  
 N(lex="now") + CONJ(lex="and") +  
 ADV(lex="then")

## 6 Problems and future work

Through an intricate system of overriding rules our grammar now prefers the analysis of time adverbial NP's and PP's as temporal adjuncts over complements. This works fine in most cases, but it often leads to loss of meaning or an incorrect analysis, as in the case of (26) below:

(26) I counted on Monday.

The sentence is bracketed by our parser as I (counted) (on Monday), incorrectly choosing the intransitive subcategorization of the verb *count*. It is our conviction that these paradoxes can only be resolved by semantic means. The grammar could be relaxed to provide both interpretations in such cases, and a semantic module would discard unlikely candidates. At the moment, however, we do not dispose of a reliable semantic module, and as our application requires us to provide the smallest possible number of parses for a sentence, we have chosen to prefer the time adverbial interpretation.

At the same time we do correctly analyze the NP as the subject of a sentence such as (27):

(27) Last year was remarkably successful.

(Hitzeman 1996) provides an intriguing array of features from a semantic perspective. Although some of these are not of much avail for a syntactic parser, his elegant set of lexical redundancy rules has prompted us to think about eliminating some of the redundancies from our flatly encoded NP's.

We have encountered adverbial NP's while encoding adverbials of time, but as (Larson 1985) points out, the same phenomenon exists for adverbials of manner and space as well, although in a much more limited way. In the light of our experiences we plan to revise the representation of these as well.

So far our rules do not cover time adverbial NP's with a clausal complement such as (28). Exploring and encoding these is among our future plans.

(28) Unfortunately, my friend arrived *the very day I left*.

As is the case with temporal NP's, the head nouns of some adverbials of space also tend to select either *in*, *on* or *at* to form a PP that answers the basic question "where." Exploring the lexical properties that govern this selection will, in our hope, enhance the precision of our grammar.

From a purely theoretical perspective, the mere existence of bare NP adverbials tempts us to rethink the traditional distinction between PP's and NP's. Hungarian being more of a head-last language, it lacks prepositions but does use postpositions. The role of an NP in a sentence is expressed by case and possibly a postposition, but it could be argued that the line between a PP and an NP is very fine if not missing altogether. For the sake of universality, it is an audacious but tempting idea to do away with this distinction in the English side of our grammar as well. Furthermore, structures such as (29) and (30) could be cited to argue for the existence of postpositions in the English language as well.

(29) My brother lived in Texas *nine years long*.

(30) He moved to California *two years ago*.

Such speculations, however, are far beyond the scope of this paper.

## References

- Bresnan, J. (2000): Lexical-Functional Syntax. Oxford: Blackwell Publishers
- Chomsky, N. (1970): Remarks on Nominalisation. In: "Readings in English Transformational Grammar" ed. R. Jacobs and P. Rosenbaum, 184-221. Waltham, Massachusetts: Blaisdell Publishing
- Corley, M. & Haywood, S. (2000): Parsing modifiers: The case of bare-NP adverbs. In: Proceedings of the 21st annual meeting of the Cognitive Science Society, 126-131. Vancouver.
- Hitzeman, J. (1996): Semantic partition and the ambiguity of sentences containing temporal adverbials *Online at* <http://www.hcrc.ed.ac.uk/publications/rp-77.ps.gz>
- Komlósy, A. (2001): A lexikai-funkcionális grammatika mondattanának alapfogalmai: Segédkönyvek a nyelvészet tanulmányozásához VII.; Kiss Gábor (ed.) Nem transzformációs nyelvtanok I. Budapest: Tinta Könyvkiadó
- Larson, R. (1985): Bare-NP Adverbs. In: Linguistic Inquiry 16, 595-621.
- McCawley, J D: (1988) Adverbial NP's: bare or clad in see-through garb? In: Language 64, 583-90.
- Prószték, Gábor & Balázs Kis (1999): A Unification-based Approach to Morpho-syntactic Parsing of Agglutinative and Other (Highly) Inflectional Languages. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, 261-268. College Park, Maryland, USA
- Radford, A. (1988): Transformational Syntax. Cambridge: Cambridge University Press