English time expressions in an HPSG grammar

Dan Flickinger
CSLI, Stanford University

1 Introduction

Syntactic analyses of time expressions in English have usually assumed that a separate idiosyncractic set of phrase structure rules would be necessary to capture the full range of facts in this domain. However, within a strongly lexicalist framework such as Head-driven Phrase Structure Grammar (Pollard & Sag, 1987, 1994), it is desirable to keep the inventory of phrasal rule schemata as general as possible, localizing idiosyncracy in the lexicon. The analysis sketched below not only maintains this generality of rule schemata, but leads to a more predictive account of these phenomena, and provides evidence that may lead to revisions in the grammar of more "central" syntactic phenomena.

The initial impetus for this study comes from work on developing an HPSG grammar of English for use in the Verbmobil machine translation project funded by the German ministry for technology research. The goal of this project is to produce a prototype by 1996 of a device which translates spoken German and Japanese into English in face-to-face dialogues. The English grammar in this prototype is used only for generation, supporting a wide range of utterance types within a constrained domain. The subject matter of the dialogues for the prototype involves two people scheduling a meeting, using a vocabulary of about 2500 words. A large number of sample dialogues have been collected to help specify the necessary vocabulary and range of grammatical coverage. The examples in (1) illustrate the kind of time expressions that appear in the English translations of these German dialogues.

(1) Dialogue extracts illustrating English time expressions

Thursday morning around nine would suit me.

Is that the eighth of April?

That's the Thursday after Easter.

In the first week of May every afternoon after two o'clock is free.

Do we want to do it right away on Monday the third of May?

Let's say two forty-five instead.

Yes, a quarter to three is okay for me too.

From ten to half past two in the afternoon I already have a meeting.

Do you mean Thursday the eighth or Thursday the fifteenth of July.

I would suggest between about ten and eleven o'clock.

The analysis of these and related expressions consists primarily of a specification of the lexical properties of the relevant words, with generalizations about these properties expressed in terms of abstract lexical types organized in a hierarchy, and related by lexical rules (cf. Flickinger (1987), Flickinger and Nerbonne (1992), Daelemanns, Smedt, and Gazdar (1992) and references cited therein). In formulating these lexical types, I assume a small set of phrasal schemata derived from those described in Pollard and Sag (1994), most of which are headed phrases with one head daughter and zero or more non-head daughters admitting a specifier, complement, or adjunct

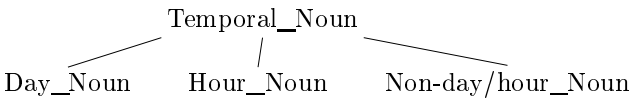
sister to the head. To employ these schemata for time expressions, then, it will be necessary to identify a unique head in each local subtree within the phrase, and to determine the relationship of the other daughters in that local subtree relative to the head.

2 Lexical Types

For this discussion I will assume without argument that the type Temporal_Noun is a subtype of Noun, which is itself indirectly one of the subtypes of Word. The precise location of Temporal_Noun within the larger lexical type hierarchy is important, but not essential to the present discussion.

The first basic division of the type Temporal_Noun, shown in (2), is motivated by facts about collocation with particular prepositions, as detailed in Quirk, Greenbaum, Leech, and Svartvik (1985). The three prepositions *on, at* and *in* restrict their arguments to three mutually exclusive subtypes of Temporal_Noun. The temporal preposition *on* requires its NP complement to be headed by a noun denoting a day, whether a day of the week, a day of the month, or a holiday. The temporal proposition *at* requires its NP complement to be headed by a noun denoting a clock time, or an event typically associated with a particular time. And the temporal preposition *in* takes as complement an NP headed by a noun denoting any of the rest of the time units, including months, years, and seasons, among others. Examples of these contrasts are given in (3).

(2)



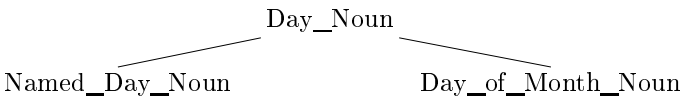
(3) a. on Tuesday, on Halloween, on the seventh
*on noon
*on October

b.*at Tuesday
at noon, at three o'clock
*at October

c.*in Tuesday
*in noon
in October, in 1995, in summer

The type Day_Noun is divided into the two subtypes shown in (4), based on the simple observation that days of the month behave syntactically like singular common nouns, with an obligatory specifier, while the other Day_Noun subtypes do not require a specifier. These contrasts are illustrated in (5):

(4)



(5) a. on Tuesday

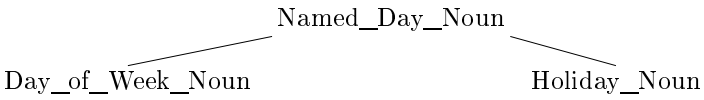
b. on the Tuesday after Halloween

c.*on seventh

d. on the seventh

The first of these subtypes is divided into the two subtypes in (6), to distinguish nouns naming days of the week, which need not be introduced by a preposition when they modify a phrase, from nouns naming holidays, which cannot directly serve as modifiers, but must appear with a preposition. This distinction is shown in (7), and is required for American English, though not for British English, in which (7a) is also ungrammatical.

(6)



(7) a. Kim will arrive Tuesday.

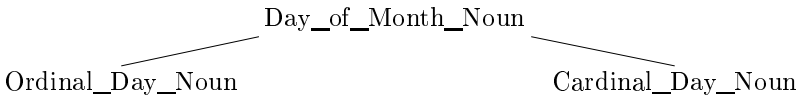
b. Kim will arrive on Tuesday.

c.*Kim will arrive Halloween.

d. Kim will arrive on Halloween.

The second subtype of Day_Noun is divided into two subtypes, as in (8), since ordinal nouns denoting days of the month can appear with either a regular determiner or a name of a month as specifier, while cardinal nouns denoting days of the month can only appear with month names as specifiers. This contrast is shown in (9).

(8)



(9) a. on the seventh of October

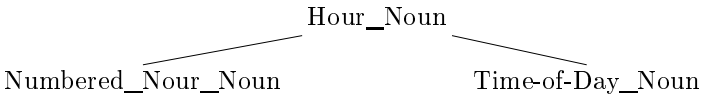
b. on October seventh

c.*on the seven of October

d. on October seven

Back at the top of this part of the type hierarchy, the second subtype of Temporal_Noun, for complements of the preposition *at*, is divided into two subtypes, shown in (10). The first is for names of the twelve hours, which can take as complement an NP denoting minutes or portions of an hour; and the second subtype is for other nouns naming conventional times of the day, such as *noon, dawn, teatime, lunch*, which cannot take such a complement. This distinction is illustrated in (11).

(10)



(11) a. at twelve fifteen

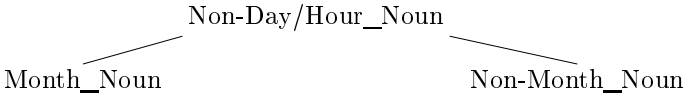
b.*at noon fifteen

c. at twelve o'clock

d.*at midnight o'clock

The third immediate subtype of Temporal_Noun can be divided into two subtypes as in (12), distinguished by the fact that nouns naming months of the year can appear as specifiers of nouns naming days of the month, unlike other temporal nouns of this subtype, as illustrated in (13).

(12)



(13) a. October seventh

b.*summer seventh (cf. the seventh day of summer)

c.*1996 seventh (cf. the seventh day of 1996)

Implicit in each of these subdivisions is a decision about which of the elements of a temporal phrase is the head, and in at least some of the above examples the choice is not transparent. I take up this issue in more detail in the final section.

Constraints on selection of complements and specifiers are represented in terms of semantic relations organized in a hierarchy. For example, Hour_Rel is a sub-type of Non_Day_Rel, which is a sub-type of Temp_Rel. While this hierarchy of relations has much in common with the lexical hierarchy of syntactic types, the two are not isomorphic, since two words may be of distinct syntactic categories, but share the same semantic relation, as illustrated in the following section.

3 Lexical Rules

For several of the subtypes introduced above, regular alternations in local structure can be observed, where a given temporal noun takes a complement or specifier optionally, or has its specifier alternate with a complement semantically, or even alternates itself between head and modifier. To capture these regularities, HPSG employs lexical rules which can be constrained to hold for particular lexical types. Two such rules are sketched in (14) and (15): the first captures the specifier/complement alternation of month nouns relative to numbered days; and the second accounts for the fact that names of months can appear as ordinary complements, but can also serve as specifiers of numbered day nouns, with a quantifier-like semantics that binds the instance variable introduced by the numbered day noun.

(14) Day_of_Month_LR:

$$\begin{array}{l}\text{Day\_of\_Month\_Noun} \\ \left[\begin{array}{ll}\text{SPR} & \langle\text{Month\_Det}\rangle \\ \text{COMPS} & \langle\,\rangle\end{array}\right]\end{array} \quad \Longleftrightarrow \quad \begin{array}{l}\text{Day\_of\_Month\_Noun} \\ \left[\begin{array}{ll}\text{SPR} & \langle\text{Det}\rangle \\ \text{COMPS} & \langle\text{Month\_Noun}\rangle\end{array}\right]\end{array}$$

(a) October seventh

(b) the seventh (of October)

(15) Month_of_Year_LR:

$$\begin{array}{l}\text{Month\_Noun} \\ \left[\begin{array}{ll}\text{CONT} & \langle\text{Month\_Rel}\rangle\end{array}\right]\end{array} \quad \Longleftrightarrow \quad \begin{array}{l}\text{Month\_Det} \\ \left[\begin{array}{ll}\text{CONT} & \left\langle\begin{array}{l}\text{Def\_Quant\_Rel}, \\ \text{Month\_Rel}\end{array}\right\rangle\end{array}\right]\end{array}$$

(a) in October

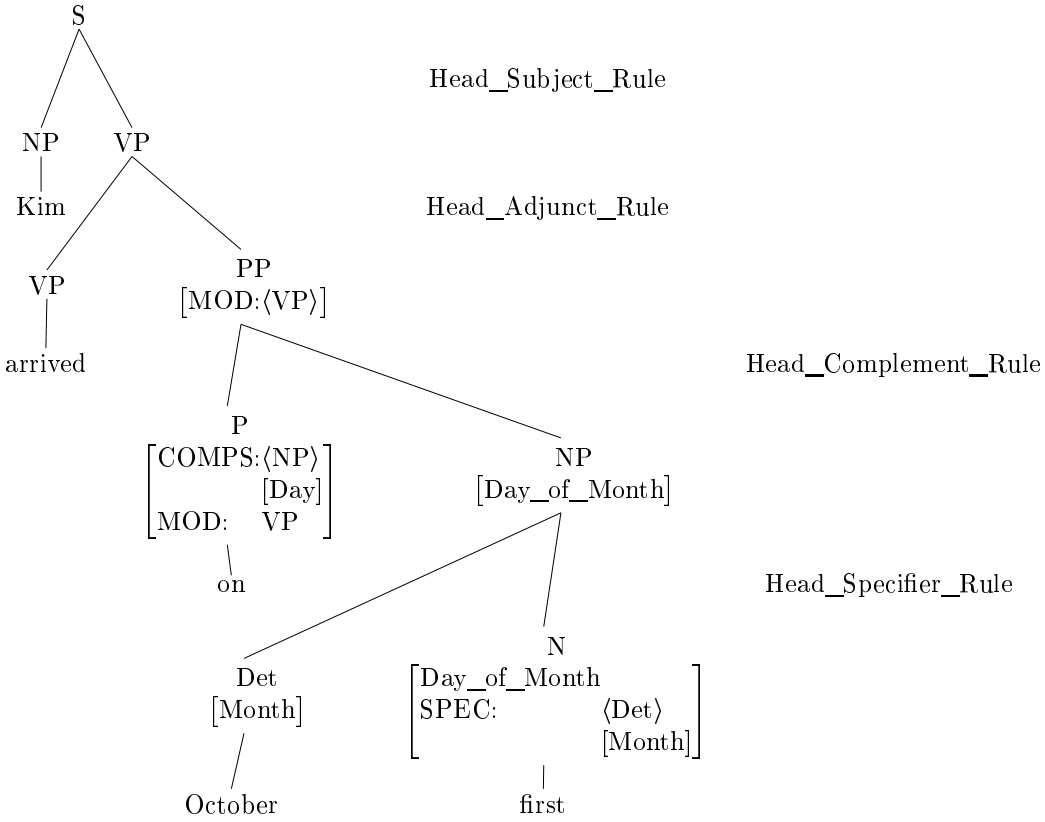
(b) on October seventh

4 Phrasal Rule Schemata

No special-purpose rules are needed, since the distributional restrictions are specified lexically as constraints on subcategorization and possible modification. By analyzing the correct lexical item as the head of the phrase, the standard syntactic machinery ensures the propagation and unification of these lexical constraints in parsing or generation. Thus the rule schemata motivated in Pollard and Sag (1994) for more conventional syntactic analyses also provide the necessary constraints on temporal phrases.

Consider the following example illustrating the application of several phrase types to provide a syntactic analysis of (16).

(16) Kim arrived on October first.



5 Some Complications

While the analysis of some temporal expressions is straightforward on the approach taken here, some systematic difficulties present themselves, and lead us to consider a wider range of data. Given our strongly lexicalist assumptions, it is not surprising that one of the first crucial steps in an analysis, and often the most difficult, is the determination of which lexical element in a temporal phrase is its head. Three examples are given here, to illustrate the nature of the problem, and to suggest strategies for resolution.

In the first example, the combination of a Day_of_Week noun and a Part_of_Day noun is a phrase that can follow the preposition *on*. The data in (17) seem to suggest that we should make *Tuesday* the head, since Day_of_Week nouns occur with *on*, while the phrase *the morning* cannot.

(17) a. on Tuesday morning

b.*on the morning

c. on Tuesday

Similarly, the phrase can serve as a VP modifier without requiring a preposition, which is a property of Day_of_Week nouns, but not of the phrase *the morning*, as seen in (18):

(18) a. Kim arrived Tuesday morning.

b.*Kim arrived the morning.

c. Kim arrived Tuesday.

However, the noun *morning* when appearing with the determiner *the* alone has different properties than when it appears with other determiners, or when it is modified, as seen in (19), where phrases headed by *morning* appear after *on* and as VP modifiers.

(19) a. on some morning

b. on the morning after Halloween

c. Kim arrived each morning.

d. Kim arrived the first morning.

Given the differences between the fixed expression *the morning* and any other noun phrase headed by *morning*, we seem to be left again without criteria to favor one element over the other as head of the phrase *Tuesday morning*. But as the data in (20) shows, there is a decisive test that leads us to treat *morning* as the head, based on the conventional assumption that inflection for number appears on the head noun in a noun phrase.

(20) a. on Tuesday mornings

b.*on Tuesdays morning

A second example of the difficulty in determining the syntactic head of a temporal phrase involves the combination of a word like *next* with a temporal noun like *October*, as shown in (21).

(21) a. Kim will arrive next October.

b. Kim will arrive next Tuesday.

c. Kim arrived last year.

Given that *next October* can modify a VP, while *October* alone cannot, an initial hypothesis is that *next* is the head of the phrase, supplying the modification properties. If *next* is a preposition heading the PP *next October*, the data in (22) would be as expected, since *on* requires an NP complement, not a PP. However, *next* will be an unusual preposition in that it is happy to combine with an unsaturated noun like *week* or *year*, though not with nouns naming time units smaller than a week nor larger than a year.

(22) a. Kim will arrive on Tuesday.

b.*Kim will arrive on next Tuesday.

c.*Kim will arrive next day.

d. Kim will arrive next week.

The third example of a construction whose head is not easily identified is illustrated in (23), consisting of a *minute* specification, a word like *before*, and an *hour* specification. The only one of the three partial phrases which can stand alone as a VP modifier is *before five*, suggesting that *before* is the head of the phrase, with *ten minutes* as a specifier, and *five* as a complement.

(23) a. Kim collapsed ten minutes before five.

b.*Kim collapsed ten minutes.

c. Kim collapsed before five.

d.*Kim collapsed five.

Similar constructions are headed by words like *to, until, of*, as shown in (24), though these phrases cannot directly modify a VP, unlike those headed by *before, after*, and these words have obligatory specifiers, not optional ones.

(24) a. Kim collapsed at ten to five.

b.*Kim collapsed ten to five.

c.*Kim collapsed at to five.

A second syntactic complication encountered in temporal phrases involves sequences of what look like temporal noun phrases. For example, in (25) we see a month noun followed by *the* followed by an ordinal (but not cardinal) day noun. Since *October* cannot serve as a VP modifier alone, while NPs headed by ordinal nouns can, it would seem that the structure of *October the third* has *third* as the head, preceded by two specifiers. This analysis is more strongly supported by the fact that the expression *October the third* appears with the preposition *on*, which selects for day nouns, rather than with the preposition *in* used with months. While the HPSG framework allows in principle for multiple specifiers on a head, they are much less commonly proposed in English than are multiple complements, making these temporal constructions especially interesting.

(25) a. Kim arrives October the third.

b.*Kim arrives October the three.

c.*Kim arrives October every third.

d. Kim arrives every third of October.

e. Kim arrives on October the third.

f.*Kim arrives in October the third.

Finally, as is to be expected in a highly grammaticized corner of the language, temporal expressions exhibit a high degree of lexical idiosyncracy, which must be accounted for in the specifications of lexical types and lexical entries, if the phrase schemata are to be left in the most general form. The first example shows idiosyncracy in a lexical paradigm, and the second shows semantic quirkiness, both of which must be represented in the lexicon, though they may not pose a real challenge for syntactic analysis. In (26) an alternation is shown between words like *last* and words like *yesterday*:

(26) a.

yesterday morning	*last morning
yesterday evening	?last evening
*yesterday night	last night
yesterday	*last day
*yestermonth	last month
?yesteryear	last year

b.

*today morning	this morning
*today evening	this evening
tonight	*this night
today	*this day
*tomonth	this month
*toyear	this year

c.

tomorrow morning	*next morning (meaning 'tomorrow')
tomorrow evening	*next evening
tomorrow night	*next night
	next month
	next year

The last example shows an idiosyncratic interpretation of *this* vs. *to-* for English, contrasted with similar expressions in German which are more regular in this respect. In (27) the combination of *this* with *morning* can be seen to be semantically bidirectional, with *this morning* referring either to the immediately preceding morning of the same day, or to the current morning. The examples in (28) show that this same bidirectionality is not present for the equivalent *tonight*, where we saw in (26) that *this* and *to-* are in some kind of alternation.

(27) a. Kim will arrive late this morning. (spoken in the morning)

b. Kim arrived late this morning. (spoken later on the same day)

(28) a. Kim will sleep well tonight. (spoken before sleeping)

b.*Kim slept well tonight. (spoken after waking the next morning)

This distinction in (28) is not found in the equivalent German expression *heute nacht*, as shown in (29):

(29) a. Kim wird heute nacht gut schlafen.
'Kim will sleep well tonight.'

b. Kim hat heute nacht gut geschlafen.
'Kim slept well last night.'

6 Conclusion

Temporal expressions, which are highly grammaticized in language, have long been recognized as a rich source of idiosyncracy for both syntactic and semantic properties. HPSG provides an appropriate framework for specifying these idiosyncracies while capturing the underlying generalizations, but I have argued here that by assuming that temporal expressions should be analyzed using the same underspecified phrasal schemata motivated for more familiar phenomena, we gain two additional benefits. First, we are led to examine further data predicted by these analyses, revealing a rich and often systematic collection of facts in this corner of the language; and second, we find evidence that sheds light on the architecture of the framework itself. This suggests that investigation of temporal expressions in other languages will be of further benefit in tuning both the particular grammars of those languages and the framework used to represent these grammars.

References

Daelemanns, W., Smedt, D., & Gazdar, G. (1992). *Computational Linguistics* 18.2: 205-218.

Flickinger, D., & Nerbonne, J. (1992). Inheritance and Complementation. *Computational Linguistics* 18.3: 269-310.

Flickinger, D. (1987). *Lexical Rules in the Hierarchical Lexicon.* Ph. D. dissertation, Stanford University.

Pollard, C., & Sag, I. A. (1987). *Information-based Syntax and Semantics, Vol. 1: Fundamentals.* No. 13 in CSLI Lecture Notes Series. Center for the Study of Language and Information, Stanford University, Stanford.

Pollard, C., & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar.* The University of Chicago Press, Chicago.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*, pp. 526-555. Longman, London and New York.