

Time Expressions in Japanese-to-English Machine Translation

Francis BOND and Kentaro OGURA

NTT Communication Science Laboratories

1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa-ken, JAPAN 239

{bond,ogura}@cslab.kecl.ntt.co.jp

Abstract

This paper describes how Japanese and English differ in expressing time, and outlines an algorithm for translating time expressions from Japanese to English. The paper deals with problems of lexical idiosyncrasy as well as the choice of articles and prepositions within time expressions. In addition time expressions are considered as parts of larger structures, and the question of whether to translate them as noun phrases or adverbials is addressed.

1 Introduction

In this paper we argue that the transfer and generation of temporal noun phrases and adverbials are best handled by a separate (though integrated) module in Japanese-to-English Machine Translation.

There are several reasons for creating a separate module. They center on the fact that temporal expressions are highly idiosyncratic, lexically and syntactically in both Japanese and English.

We argue that attempting to handle all phenomena by a single mechanism, while conceptually elegant, causes unnecessary complications. Instead we adopt the multi-level machine translation approach of Ikehara *et al.* (1991) in which there are many levels of transfer between two languages, and expressions are transferred at whatever level the system judges to be the most appropriate.

The processing described has been implemented in the Japanese-to-English machine translation system ALT-J/E (Ikehara *et al.* 1996).

2 Temporal Noun Phrases

English temporal noun phrases, along with locatives, do not have all the properties of prototypical 'purebred' noun phrases. They are one of the five types of noun phrases classed as 'defective' by Ross (1995). For example, in contrast with 'purebred' noun phrases, they are typically pronominalized by *then* rather than *it*. There is also a great deal of variation in realization between dialects, with substantial differences between Australian, American and British English. In addition,

there is a lot of lexical idiosyncrasy as we will show in the following sections.

In the following section we will restrict our discussion to noun phrases that show position in time, rather than duration or frequency. English temporal adjuncts are described in some detail in Quirk *et al.* (1985:526-555). To the best of our knowledge, the only description of their treatment in natural language processing is that of Flickinger (1996), who discusses English time expressions in the HPSG grammar used by the Verb-mobil German/Japanese-to-English Machine Translation project.

2.1 Temporal Noun Phrase Structure

English time position noun phrases used primarily to refer to time (for example within adverbials or the subject of sentences such as *Spring has come*) are highly idiosyncratic in their lexical choice, as well as their choice of determiners, typically having no surface determiner, although some take the definite article. Note that noun phrases headed by the same nouns, but not primarily referring to time, behave as do other nouns: *It was a spring to remember*. To handle these lexical and syntactic idiosyncrasies we introduce special processing for temporal noun phrases.

Ross (1995:433) claims that such defective noun phrases are always locally triggered, that is there is some clause mate that forces or enables the noun phrase to become defective. If we allow time position adjuncts to license themselves as being defective, then this claim holds, and we can always count on there being some trigger to introduce our special processing.

We will now give some example of the idiosyncrasies, starting with the lack of an article for unmodified nouns, assuming they are in temporal noun phrases.¹

- | | | |
|-----|-------------------------------|-----------------|
| (1) | a. today, yesterday, tomorrow | (deictic-day) |
| | b. Monday | (day-of-week) |
| | c. Christmas | (holiday) |
| | d. 3 o'clock, 12:15 | (numbered-hour) |
| | e. dawn | (time-of-day) |
| | f. winter | (season) |

¹After each example in (1) we give the semantic attribute of its head, from the hierarchy given in Figure 2.

We analyze the noun phrases with no surface determiner as NULL determiners, phonologically empty determiners with definite reference.

In contrast with the noun types shown in (1), ordinal numbers denoting the day of a month, on the other hand, normally take the definite article: *the 19th*.

The choice of modifier also affects the choice of determiner, for example, there are some temporal expressions which take the null determiner when they are modified by another class of time expressions (2):

- (2) a. Monday morning
- b. yesterday morning
- c. Monday night
- d. February 19
- e. February 19th

We analyze these simply as “NULL modifier head”. In our translation module, we generate the determiner during the transfer from Japanese. Flickinger (1996) introduces a more elegant analysis where the first element is treated as specifier (determiner) of the second, thus explaining the lack of article. Both the specifier’s change in part of speech, and the head’s choice of complement, are expressed by lexical rules. However, as null determiners or their equivalent are needed for the noun phrases in (1) anyway, we do not consider their use here to be problematic.

In addition, as Flickinger (1996) notes, his analysis leads to noun phrases with multiple specifiers in expressions like *February the 19th*, although most analyses of English allow only one specifier.

To handle such cases, we use a special structure, the special compound noun phrase, that we have established for noun phrases where there is no obvious head, such as person and company names, and addresses. We use this structure for time position noun phrases which include the following elements: year, day-of-month, month and/or numbered-time. In these noun phrases there is no obvious semantic head and there are many possible representations in English. We show some examples of the choice of expressions for a single date in (3). The choice of representation is mainly a question of style. In particular, it does depend on the Japanese source noun phrase, which has only two possible forms: *2-gatsu-19-nichi* “2 month 19 day” and *2-gatsu-no-19-nichi* “2 month GEN 19 day”.

- (3) February the 19th *vs* February 19 *vs* February 19th *vs* the 19th of February

By establishing a set of special structures for noun phrases that behave atypically, and thus have to be treated atypically anyway, we are able to preserve a uniform structure for all other noun phrases (with a single, although potentially phonologically empty,

specifier). Although the grammar is consistent with ‘purebred’ noun phrases, the choice of determiner is not, which is why we argue for a separate module.

In the next section we will describe the transfer module for temporal noun phrases, which does most of the hard work in creating appropriate English structures, handling lexical and phrasal idiosyncrasy.

2.2 Transfer and Generation

Temporal noun phrases in Japanese can be considered to be of three types; those headed by single nouns, those headed by compound nouns, and those made of one temporal noun phrase modifying another.

The transfer stage for noun phrases headed by single nouns is basically a process of replacing them by their equivalents in the lexicon, which may be a single English noun: *kinō* “yesterday” or a phrase: *ototoi* “the day before yesterday”.

Dates (years, months, days of months, and numbered times) are compound nouns in Japanese, typically consisting of a number and a temporal noun. The compound noun rules first distinguish between time positions and time periods, for example, the adverbial *13-nichi* “13 days” could be *on the thirteenth* or *for thirteen days*. Once it has been determined that the noun phrase refers to time position, simple regular rules are used to generate the corresponding English expressions. For example, days of months in Japanese have the form NUMERAL-*nichi* “NUMERAL day” and are translated into special compound noun phrases with the day-of-month slot filled by the value for the numeral.²

Complex noun phrases require more complicated rules. We show the rules for the combination of a noun phrase denoting deictic-day or day with one denoting period-of-day (*morning*, *afternoon*, *evening*) or *night* in Figure 1. The Japanese will be of the form α -*no* β “ α -GEN β ” where α is headed by a day or deictic-day noun and β is headed by a period-of-day noun or *night*, that translates into English noun phrase B.

These transfer rules capture the lexical and phrasal idiosyncrasies of the temporal noun phrases, and are relatively easy to test and expand. The rules are all specific to temporal noun phrases, no generality has been lost by putting them in a separate module.

Note that these rules do not necessarily preserve the Japanese structure. In particular many temporal expressions made up of two noun phrases in Japanese are

²This is further complicated by the fact that the Japanese have two counting systems for years, one based on the western system (A.D.) and the other on years of the current emperor’s reign. *3-nen* “3 years” is thus multiply ambiguous between, at least, 3 AD, 1903 AD, 2003 AD, 1991 AD (the third year of the current emperor’s reign) and a three year period.

B is period-of-day or night:
if A is deictic-day
 if A = *ototoi* "the day before yesterday"
 ⇒ *the B before last* ; 2 before
 if A = *kinō* "yesterday"
 if B = *night*
 ⇒ *last night*
 else
 ⇒ NULL yesterday B ; 1 before
 if A = *kyō* "today" or *honjitsu* "today"
 if B = *night*
 ⇒ NULL tonight
 else
 ⇒ *this B* ; today's
 if A = *ashita* "tomorrow"
 ⇒ NULL tomorrow B ; 1 after
 if A = *asatte* "the day after tomorrow"
 ⇒ *the B after next* ; 2 after
 if A = *shiasatte* "the day after the day after tomorrow"
 ⇒ *the B after the B after next* ; 3 after
if A is relative-day
 if A = *zenjitsu* "the previous day"
 ⇒ *the previous B*
 if A = *yokujitsu* "the following day"
 ⇒ *the following B*
if A is named-day
 ⇒ NULL A B
if A is day-of-month
 ⇒ *the B of the A* (A must be ordinal)

Figure 1: Rules for day, deictic-day and period-of-day

often most naturally translated as a noun phrase and an adverbial phrase in English. For example, *ashita-no akegata* "tomorrow-GEN dawn" could possibly be translated as *tomorrow's dawn*, but is generally translated as *dawn tomorrow* where *tomorrow* is an adverbial modifying *dawn*.

Attempts to preserve structure, translating all Japanese temporal noun phrases as English noun phrases, are problematic as the resulting expressions are often unwieldy or ambiguous. Consider for example *kotoshi-no kurisumasu* "this year-GEN Christmas", which could potentially be translated as *this year's Christmas*, *this Christmas* or *Christmas this year*. *this year's Christmas* sounds extremely unnatural. The use of deictic modifiers like *this* and *last*, are not bound by year boundaries, *this Christmas* spoken in January, could refer to Christmas last year, or the coming Christmas and is thus ambiguous in a way that the Japanese original is not. The only acceptable translation of *kotoshi-no kurisumasu* "this year-GEN Christmas" is *Christmas this year*, where *this year* is an adverbial.

We also found, in a preliminary investigation of bilingual corpora, that translation of Japanese tem-

poral noun phrases into English adverbials was very common, not only in complex temporal noun phrases, but also in noun phrases with only one part being temporal, for example those in (4).

- (4) *konshū-no uchiawase* "this week-GEN meeting"
⇒ *this week's meeting* vs *the meeting this week*

At present our rules translate most Japanese complex temporal noun phrases, and all temporal noun phrases occurring inside larger non-temporal noun phrases as noun phrases in English, but as we identify more criteria for choosing between adverbials and noun phrases they will be added to our module.

Note that we have similar sets of rules to those presented in Figure 1 for other time combinations, such as: *the January before last* (year + month), *next Saturday* (week + day), *Christmas morning* (holiday + period-of-day) and so on.

3 Temporal Adverbials

As well as translating Japanese adverbials as adverbials it is often necessary to translate Japanese temporal noun phrases into English adverbials, as discussed in Section 2.2. In both cases, we need to select which preposition to use. The main problem for generating English prepositions is deciding between *on*, *at* and *in*. Other prepositions, such as *before* or *during* have Japanese equivalents (normally functional nouns or post-positional particles) which can be used to select them. The most widely used particle in Japanese for temporal expressions is *ni*, which can be translated as *on*, *at* or *in*. To choose between these three alternatives, we need to consider both the semantic attribute of the head noun, and its modifiers, in English. We use the semantic attribute system for Japanese analysis of Ikehara *et al.* (1991), augmented with several types defined especially in English generation. The relevant parts of the hierarchy are shown in Figure 2. The nodes underlined are those that appear in the Japanese semantic attribute system.

Our type hierarchy matches that suggested by Flickinger (1996), and we adopt his nomenclature. The hierarchy presented here is very similar, but includes deictic-day nouns and more branches under the day and non-day/hour node. The new nodes are needed in writing the rules to translate complex temporal noun phrases outlined in Section 2.2, and in choosing prepositions, as will be shown below.

The hierarchy is loosely based on the choice of preposition for temporal adverbials formed by a prepositional phrase followed by an unmodified noun phrase headed by each type. deictic-day nouns take no preposition, they form adverbials by themselves.

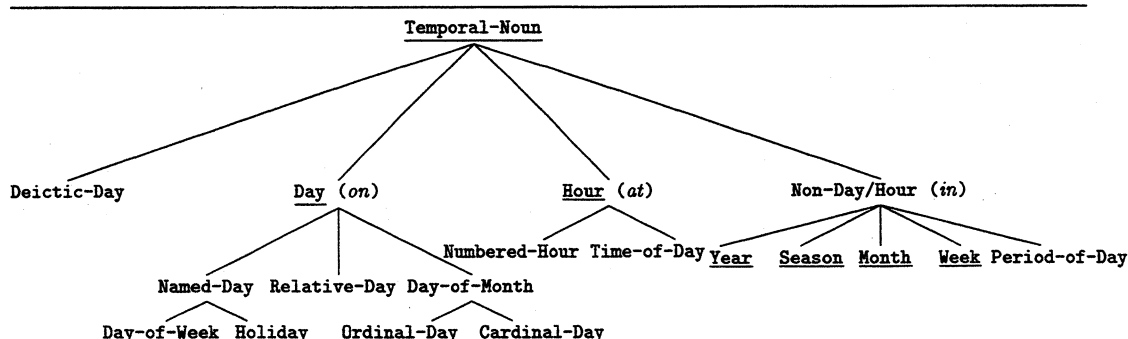


Figure 2: Noun Type Hierarchy

Noun phrases headed by day nouns will take the preposition *on*. The preposition is optional for noun phrases headed by day-of-week nouns in American English. Noun phrases headed by day-of-month nouns can take months as specifiers in Flickinger's analysis, cardinal-day nouns can only take months, while ordinal-day nouns can take months or *the* or both.

Noun phrases headed by hour nouns, including numbered-hour nouns and time-of-day nouns like *noon* and *dawn* are selected by the preposition *at*. The common denominator seems to be the idea of a precise moment in time.

The remainder of noun phrases headed by temporal nouns are selected by the preposition *in*.

Our algorithm for choosing whether to use a preposition, and if so which preposition to use, is given in Figure 3. The algorithm is used to convert noun phrases headed by temporal nouns to adverbials, and has been extended to include noun phrases of the form *A of B* where *A* is *beginning*, *middle*, *end* and *B* is a temporal noun phrase. There are also some event nouns that could be analyzed as heading temporal noun phrases that can appear in temporal adverbials, such as *opening* in *at the opening [of the Osaka Stock Exchange]*, we are extending our analysis to include them.

on is optional for some temporal expressions. Whether to use it depends on dialect, register, domain and genre (e.g. stock reports and speech tend to omit *on*). We always generate *on*, under the assumption that it is easier to delete it afterwards than put it in.

4 Conclusion

In this paper we present an outline of rules for translating Japanese temporal phrases to English. Due to the unproductive nature of the rules, and the many lexical idiosyncrasies, we argue that these rules are better

```

If equal to deictic-day or tonight
or (premodified by a deictic term this, that, last, next
or one of the deictic-day nouns
or a quantifier such as every, some)
or post modified by ago, later
⇒ adverbial is a noun phrase (no preposition)
else create a prepositional phrase headed by:
  If head is (unmodified (hour or night)) or beginning, end:
    ⇒ at
  Else-If head is day or (period-of-day and modified)
    ⇒ on
  Else
    ⇒ in
  
```

Figure 3: Preposition choice for temporal adverbials

thought of as a separate module, although integrated with the entire system.

References

- FLICKINGER, DAN. 1996. English time expressions in an HPSG grammar. In *Studies on the Universality of Constraint-Based Phrase Structure Grammars*, Technical Report 06044133, 1-8. Osaka University, Japan.
- IKEHARA, SATORU, SATOSHI SHIRAI, and FRANCIS BOND. 1996. Approaches to disambiguation in ALT-J/E. In *Proceedings of the International Seminar on Multimodal Interactive Disambiguation (MIDDIM '96)*, 107-117.
- IKEHARA, SATORU, SATOSHI SHIRAI, AKIO YOKOO, and HIROMI NAKAIWA. 1991. Toward an MT system without pre-editing - effects of new methods in ALT-J/E-. In *Proceedings of MT Summit III*, 101-106. (cmp-lg/9510008).
- QUIRK, RANDOLPH, SIDNEY GREENBAUM, GEOFFREY LEECH, and JAN SVARTVIK. 1985. *A Comprehensive Grammar of the English Language*. Essex: Longman.
- ROSS, JOHN ROBERT. 1995. Defective noun phrases. In *CLS 31*, 398-440.