A Language-Neutral Representation of Temporal Information

Richard Campbell*, Takako Aikawa, Zixin Jiang, Carmen Lozano, Maite Melero and Andi Wu

Microsoft Research
One Microsoft Way, Redmond, WA 98052 USA
{richcamp, takakoa, jiangz, clozano, maitem, andiwu}@microsoft.com
*to whom inquiries should be addressed

Abstract

We propose a framework for representing semantic tense that is language-neutral, in the sense that it represents what is expressed by different tenses in different languages in a shared formal vocabulary. The proposed framework allows the representation to retain surface distinctions for particular languages, while allowing fully semantic representations, such as a representation of event sequence, to be derived from it. The proposed framework also supports the incorporation of semantic tense information that does not derive from grammatical tense, but derives instead from other expressions such as time adverbials. The framework is currently implemented in NLPWin, a multi-lingual, multi-application natural language understanding system currently under development at Microsoft Research, but the representational framework is in principle independent of any particular system.

1. Introduction[1]

Multilingual applications face (at least) two problems in the domain of semantic tense: First, there is the problem that grammatical, or morphological, tenses in different languages do not mean the same thing. In English, for example, grammatical past tense situates an event prior to the utterance ("speech time" in Reichenbach's (1947) terminology), and grammatical present tense situates an event simultaneous with the utterance. In contrast Japanese past tense situates an event prior to some reference time, and present tense situates an event simultaneous with some reference time, where the reference time may or may not be the utterance time. Neither language has a tense that expresses exactly what is expressed by past or present in the other. This poses a problem for applications such as machine translation (MT), since a given grammatical tense in one language does not automatically translate into the same surface tense in another language:

(1) 彼女は病気だと言った。
kanozyo-wa [byooki **da**] to itta
she -Top sick be**-Pres** that say-Past
'she said [she **was** sick]'

In (1), for example, the grammatical present tense in the embedded clause (indicated by brackets) translates into English as grammatical past tense, both of which allow the interpretation that the event described in the embedded clause is simultaneous with that described in the main clause.

Another problem is that what is expressed as grammatical tense in one language is sometimes only expressible as an adverbial construction in another language. For example, Chinese has no grammatical tense per se (see Section 3.3 for more details); consequently a single form can in principle express past, present or future; this is illustrated in the following examples:

(2) 昨天他来看我
zuotian ta lai kan wo
yesterday he come see me
'Yesterday he came to see me.'

(3) 明天他来看我
mingtian ta lai kan wo
tomorrow he come see me
'Tomorrow he will come to see me.'

In (2) and (3), the adverbials *zuotian* 'yesterday' and *mingtian* 'tomorrow' are all that indicate that these sentences are set in the past and future, respectively.

In this paper, we propose a framework for representing semantic tense, by which we mean information about the sequence of events. Our framework is *language-neutral*, in the sense that it represents surface tense marking of various languages using a shared formal vocabulary. Our framework also allows the incorporation of semantic tense information that is not expressed as grammatical tense, for example, that (2) is about a past time. Also, since a large part of what is expressed by tenses concerns the sequence of events and states, one aspect of our framework is enabling an explicit representation of temporal sequence. The analyses reported here are currently implemented in the NLPWin system under development at Microsoft Research (Heidorn, 2000).

Most (if not all) other proposals for a language-neutral representation of tense, such as Van Eynde (1997), are explicit attempts to represent the semantics of tense directly. However, the kind of semantic representation of tense may vary considerably depending on application. For example, some applications may require tense to be represented in first-order predicate calculus, perhaps incorporating Davisonian event arguments (Davidson, 1980), while others might require only an explicit sequence of events, as in Filatova and Hovy (2001).

The novelty of our approach lies in the fact that it does not attempt to be a particular semantic representation. Our goal is to preserve syntactic information about semantic tense so that various semantic representations of

[1] We would like to thank three anonymous reviewers and our colleagues in the Natural Language Processing group at MSR for their helpful comments and discussion, especially Michael Gamon, Marisa Jimenez, Jessie Pinkham and Hisami Suzuki.

tense can be constructed if necessary for a particular application. For example, our representation is compatible with both the referential theory of tense (e.g. Enç, 1987) and the quantificational theory of tense (e.g. Ogihara, 1995). Also, although it does not express sequence of events directly, a representation of such a sequence can be derived from our representation.

Our framework owes much to Reichenbach (1947); but while a strictly Reichenbachian approach to tense may work well for European languages, such an approach becomes unwieldy when faced with a set of languages with more typologically diverse tense systems, including Japanese and Chinese, aspects of which are discussed below. We therefore do not rely on the Reichenbachian notions of reference and event times, as does e.g. Van Eynde (1997), but adapt what we take to be Reichenbach's essential insights to a wider range of tense systems.[2]

Before proceeding, it is necessary to say something about the terms *tense* and *aspect*, and to lay out what the scope of the paper is. By *semantic tense*, we mean information about how events or situations are sequenced; this includes some of what in some traditions is called aspect, such as the interpretation of the English perfect, etc. It also includes information that may not be recorded by grammatical tense, as shown in (2) and (3). By *aspect*, we mean temporal information that goes beyond temporal sequence, such as (im)perfectivity, progressive, stative, habitual, and the like. In this paper, we are concerned with semantic tense, not primarily with aspect, though some aspectual features are considered in Section 3.3.2, below.

The paper is organized as follows: In Section 2 we outline the general framework of Language-Neutral Syntax (LNS) (Campbell, 2002; Campbell & Suzuki, 2002), within which we situate the current proposal; in Section 3, we lay out our proposal for the representation of semantic tense; in Section 4, we compare our system to other proposals for representing semantic tense; Section 5 offers a conclusion.

2. Language-neutral syntax

In this section we describe the basic properties and motivation for LNS. For more detailed descriptions, the reader is referred to Campbell (2002) and Campbell & Suzuki (2002).

LNS is a level of representation that is more abstract than a surface-syntactic analysis, yet not as abstract as a fully-articulated semantic analysis; rather, it is intermediate between the two. The basic design principle of LNS is that it be close enough to the surface syntax of individual languages to allow reconstruction of the surface structure of a given sentence (i.e., LNS can serve as the input to a language-particular generation function), yet abstract and language-independent enough to allow derivation of deeper semantic representations, where necessary, by a language-independent function. The role of LNS is illustrated schematically in Figure 1.

[2] However, we do use the terms "reference time" and "event time" informally below.

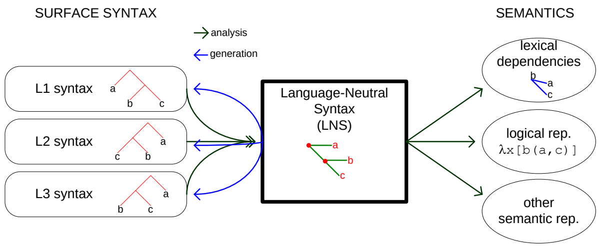

Figure 1: Language-Neutral Syntax

The primary motivation for such an intermediate representation is to mediate between languages in multilingual applications, given that fully articulated semantic representations are typically not needed in most such applications. For example, the Adjective + Noun combinations *black cat* and *legal problem* have identical surface structures, but very different semantics: the first is interpreted as λx[black(x) & cat(x)], i.e., as describing anything that is both a cat and black; the second, however, does not have the parallel interpretation as a description of something which is both a problem and legal: rather, it typically describes a problem having to do with the law. To accurately analyze this distinction would require extensive and detailed lexical annotation for adjective senses and, most likely, for lexicalized meanings of particular Adj + Noun combinations; such extensive annotation, if it is even possible, would make a system that depends on it very brittle. For most applications, however, this semantic difference is immaterial, and the extensive and brittle annotation unnecessary: for example, all that we need to know to translate these phrases into French *chat noir* lit. 'cat black' and *probléme legal* lit. 'problem legal' is that the adjective modifies the noun in some way. LNS is a representation in which *black cat* and *legal problem* have the same structure, despite their deep semantic difference, and in which *black cat* and *chat noir* have the same structure, despite their superficial syntactic difference.

An LNS representation is an annotated tree, in which constituents are unordered, and linked to their parent by labeled arcs, the labels corresponding to semantically motivated grammatical functions such as semantic head, logical subject, time, etc. The LNS tree is annotated with semantically motivated features and relations expressing long-distance dependencies (such as binding and control) and discourse-oriented functions (such as topic and focus). An example (somewhat simplified, and with tense not represented for the time being) is given below; this figure represents the LNS for this noun phrase before the implementation of the framework for tense representation presented below.

(4) the cat that was seen yesterday

```
NOMINAL1 (+Def +Sing)
|_SemHeads--cat1
|_L_Attrib--FORMULA1 (+Pass +Proposition)
            |_SemHeads--see1
            |_L_Sub---_X1
            |_L_Obj---NOMINAL2
                      |_SemHeads--that1
                      |_Cntrlr:  cat1
            |_L_Time-- yesterday1
```

The root node (NOMINAL1) is in the upper left; the daughters of a given node are indicated by labeled arcs such as *SemHeads* (semantic head), *L_Attrib* (logical attributive modifier), *L_Obj* (logical object), and the like. In addition to these attributes indicating deep grammatical relatons, there are other attributes which express additional relations among nodes in the tree. For example, the relative pronoun NOMINAL2 has a *Cntrlr* attribute, whose value is *cat1*, and indicates that *cat1* is the antecedent of the relative pronoun. The Cntrlr attribute is not part of the LNS tree per se; that is, the value of Cntrlr must be part of the LNS tree independently of the Cntrlr relation (in this case, as the semantic head of NOMINAL1). We refer to attributes such as Cntrlr as *non-tree* attributes. For display purposes only in this paper, we display non-tree attributes as labeled arcs, even though they are not part of the LNS tree per se; they will be displayed slightly differently, however, in that the value of the attribute is introduced by a colon, instead of by a dashed line.

In this example we see also that passives are normalized in terms of their argument structure, but the fact that the relative clause is passive is recorded in the feature +Pass on FORMULA1. This reflects a basic design principle of LNS: The basic structure is normalized for variation both within and among languages, but surface distinctions (such as the active/passive distinction) are retained as much as possible.

Thus an LNS representation needs to be close enough to the surface syntax to indicate meaningful distinctions, yet abstract enough to normalize meaningless cross-linguistic variation.

3. Framework for semantic tense

The LNS representation of semantic tense must therefore satisfy the following design criteria:

(5) *Design criteria for LNS representation of tense*:
a. Each individual grammatical tense in each language is recoverable from LNS.
b. The explicit sequence of events entailed by a sentence is recoverable from LNS by a language-independent function.

Criterion (5)(a) says that we must be able to reconstruct, by a distinct generation function for each language, how the semantic tense was expressed in the surface form of that language; this criterion will be satisfied if the LNS representation is different for each tense in a particular language. Criterion (5)(b) says that we must be able to derive an explicit representation of the sequence of events from an LNS representation by means of a language-independent function. This criterion will be satisfied if the representation of each tense in each language is truly language-neutral. In this section we detail a framework for semantic tense that meets the design criteria in (5). We begin by giving the details of the basic formalism (which we will add to in subsequent subsections), followed by a discussion of the motivation and function of its various aspects.

3.1. Basic framework: simple tenses

3.1.1. Tense features and relations

In our proposal each tensed clause contains a distinct Tense node, which is in the *L_Tense* ("logical tense") relation with the clause, and which is specified with semantic tense features, representing the meaning of each particular tense, and attributes indicating its relation to other nodes (including other Tense nodes) in the LNS tree. Semantic tense features can be either *global* or *anchorable*.[3]

The basic tense features, along with their interpretations, are given in the following tables; Table I shows the global features, and Table II the anchorable ones ('U' stands for the utterance time: 'speech time' in Reichenbachian terms):[4]

Feature	Meaning
G_Past	before U[5]
G_NonPast	not before U
G_Future	after U

Table I: Global tense features

Feature	Meaning
Befor	before Anchr if there is one; otherwise before U
NonBefor	not before Anchr if there is one; otherwise not before U
Aftr	after Anchr if there is one; otherwise after U
NonAftr	not after Anchr if there is one; otherwise not after U

Table II: Anchorable tense features

The tense features of a given Tense node are determined on a language-particular basis according to the interpretation of individual grammatical tenses. For example, the simple past tense in English is [+G_Past], the simple present is [+G_NonPast +NonBefor], etc.

Additional features may turn out, on further analysis, to be necessary; for example, many languages make a grammatical distinction between immediate future and general future, or between recent past and remote or

[3] The distinction between global and anchorable tense features is very similar to Comrie's (1985) distinction between 'absolute' and 'relative' tenses. We have adopted the different terminology to emphasize that the global/anchorable distinction is for features, not for tenses per se, as in Comrie's taxonomy.

[4] Note that, given their meanings, some pairs of Tense features are semantically incompatible with each other, and cannot occur on the same node. For example, a given Tense cannot be [+G_Past +G_NonPast].

[5] Strictly speaking the meaning of the global tense features is to express a relation between a given time t and a globally specified reference time, G. Conceivably, the value of G could vary, depending on various factors including genre, discourse context, etc. However, we currently have no theory as to how G might be set to any value other than U, so we will assume throughout that the global referene time is always the same as the utterance time.

general past. We have nothing to say about these specific contrasts, however, other than to note that the framework we propose is flexible enough to accommodate new tense features, if necessary.

A Tense node T will also under certain conditions have a non-tree attribute called *Anchr*, which indicates a relation that T bears to some other Tense node (the value of the Anchr attribute must be another Tense node). Like other non-tree attributes such as Cntrlr, Anchr should be thought of as an annotation on the basic tree, not as part of the tree itself; that is, the value of the Anchr attribute must fit into the LNS tree in some independent way. A Tense node has an Anchr attribute if (a) it has anchorable tense features; and (b) meets certain structural conditions. For simple tenses, the structural condition that it must meet to have an Anchr is that the clause containing it is an argument (i.e., logical subject or object) of another clause; in this case the value of Anchr is the Tense node in the governing clause. In the discussion of compound tenses below we will augment the set of sufficient structural conditions for having an Anchr.[6]

3.1.2. Past tense in English and Japanese

As indicated in Table II, if a Tense node with anchorable features has no Anchr, then it is interpreted as if anchored to the utterance time U. This means that, for example, a [+G_Past] Tense and an unanchored [+Befor] Tense have the same interpretation, all else being equal. Consider the following English and Japanese sentences, with the relevant parts of their LNS structure shown:[7]

(6) She was sick.
FORMULA1
|_SemHeads----sick1
|_L_Tense----_Tense1 (**+G_Past**)

(7) 彼女は病気だった。
kanozyo-wa byooki datta
she -Top sick be-Past
'She was sick.'
FORMULA1
|_SemHeads----病気1 (sick)
|_L_Tense--_Tense1 (**+Befor**)

The English and Japanese past tenses are represented differently because they are semantically different, though in these simple examples that difference is neutralized. The English simple past tense is [+G_Past], indicating that it denotes a time that is before U. The Japanese simple past tense on the other hand is [+Befor], indicating that it denotes a time that is before its Anchr. However, in this simple root sentence, there is no Anchr, so it is interpreted as if anchored to U; hence the interpretation is before U. Thus the design criterion (5)(b) is met, at least for these simple cases: a simple language-independent function would yield the correct sequence *be_sick* < *U* for both these examples.

The semantic difference between the English and Japanese past tenses comes into play when the Anchr attribute is present, which for simple tenses is in clauses that are arguments of a higher clause. Consider the following English and Japanese examples, in which the tense in question (in boldface) is in an embedded sentence (indirect speech), represented in LNS as the logical object (L_Obj) of the matrix clause:

(8) She said she was sick.
FORMULA1
|_SemHeads--say1
|_L_Tense--_Tense1 (+G_Past)
|_L_Obj--FORMULA2
    |_SemHeads--sick1
    |_L_Tense--**_Tense2** (+G_Past)

(9) 彼女は病気だったと言った。
kanozyo-wa byooki datta to itta
she -Top sick be-Past that say-Past
'she said she was sick'
FORMULA1
|_SemHeads--言う1 (say)
|_L_Tense--_Tense1 (+Befor)
|_L_Obj--FORMULA2
    |_SemHeads--病気1 (sick)
    |_L_Tense--**_Tense2** (+Befor)
        |_Anchr: _Tense1

Since the embedded tense in (8) is +G_Past, its interpretation is before U; left unspecified is whether the situation described by the embedded clause (FORMULA2) is reported to have occurred before, or simultaneous with, the situation described by the matrix clause. In fact, both interpretations are possible in this case: her reported sickness may be simultaneous with her saying that she was sick (i.e., she said "I am sick"), or it may have preceded it (i.e., she said "I was sick").[8] The structure we assign to it captures that underspecification succinctly.

In (9), on the other hand, the embedded tense, *_Tense2*, is +Befor; since it has an anchorable feature, and its clause is the logical object of another clause, it must be anchored to the tense of that matrix clause, i.e., to *_Tense1*. Consequently, it denotes a time that is before the time denoted by *_Tense1* (which, like *_Tense1* in (7), denotes a time before U). So the only interpretation (9) has is that her reported sickness is prior to her saying that she was sick; i.e., it can only mean 'she said "I was sick"'; it cannot mean 'she said "I am sick"'. This construction illustrates the essential difference between the English and Japanese past tense forms: the former directly expresses a

[6] We have not ruled out the possibility of language-particular anchoring conditions, but so far have not encountered any need for them.

[7] In this paper we show only the parts of the LNS necessary to illustrate the treatment of tense; for example, we leave out logical subject, etc., unless otherwise necessary. Note also that the copula is regularly omitted from the LNS (see Campbell, 2002).

[8] A third logical possibility, consistent with the interpretation of G_Past, is that her sickness was in the past (i.e., before U), but after her saying that she was sick; i.e., she said "I will be sick". But this kind of interpretation seems to be universally disallowed without some kind of irrealis marking on the clause (such as a modal), and therefore does not need to be separately indicated.

relation to U, while the latter directly expresses a relation to some "reference" time, which may or may not be U.

Examples such as (8) and (9) illustrate precisely why the English and Japanese grammatical past tenses have different representations in the current framework. Suppose for example that the Japanese past tense were [+G_Past] (like the English past), instead of [+Befor]; then Japanese (9) should have the same range of interpretations as English (8), in particular it should be able to serve as a description of an event in which she said "I am sick"—i.e., where the time of her being sick coincides with the time that she said she was sick. As noted, however, this interpretation is not available for (9), as it is for (8).

Our analysis of the English and Japanese past tenses differs from the approach taken by e.g. Ogihara (1995), who claims that English and Japanese past tenses mean the same thing, and that differences such as that between (8) and (9) below are due to the optional application in English of a rule that deletes the embedded past tense from the logical form component. Our analysis gives a uniform description to both the English and Japanese grammatical past tenses.

It is important to note that there is only one sense of the feature Befor (the same holds true for all the anchorable features in Table II), and hence only one meaning for Japanese past tense, in our system. This is a crucial point which is easily overlooked: phrased in strictly Reichenbachian terms, we may appear to be saying that the Japanese past tense means *either* E<R (if it is anchored) *or* E<S (if not anchored). But this appearance of bi-vocalism is due, we believe, to an overly rigid adherence to Reichenbach's notation; our own notation is more flexible, allowing us to characterize the Japanese past tense as univocal, while still retaining what we regard as Reichenbach's essential insights, namely that some tenses relate to U and others to a structurally determined "reference" time.

3.1.3. Present tense in English and Japanese

Another good illustration of the differences between global and anchorable tense features is provided by the English and Japanese present tenses. As in the case of past tense, the two tenses receive the same interpretation in simple sentences:

(10) She is sick.
FORMULA1
|_SemHeads—sick1
|_L_Tense--_Tense1 (**+G_NonPast** +NonBefor)

(11) 彼女は病気だ。
kanozyo-wa byooki da
she-Top sick be-Pres
'She is sick'
FORMULA1
|_SemHeads—病気1 (sick)
|_L_Tense--_Tense1 (**+NonBefor**)

Since the English present tense in (10) is [+G_NonPast] (as well as [+NonBefor]; see below), it must denote a time that is not before U. The Japanese present tense is just [+NonBefor], so it denotes a time that is not before its Anchr; since it lacks an Anchr, in this case, it must denote a time that is not before U. Consequently (10) and (11) receive the same interpretation.

Note that nothing in these representations directly expresses anything about the "present": G_NonPast is interpreted as "not before" U, but does not have to be simultaneous with U. This is by design: the English grammatical present tense allows a future interpretation as well as a "present" one, as in *We speak tomorrow* (see Section 4, below). Our assumption is that present-time reference is the default denotation for any Tense whose features and relations to other time expressions are consistent with that interpretation. Similar comments hold for the Japanese present tense, which is [+NonBefor] in our analysis. As in English, the Japanese present tense also allows a future-time construal (see Section 3.3.3, below).

As in the case of the past tenses, the difference between the English and Japanese present tenses shows up when there is an Anchr:

(12) She said she is sick.
FORMULA1
|_SemHeads--say1
|_L_Tense--_Tense1 (+G_Past)
|_L_Obj--FORMULA2
|_SemHeads--sick1
|_L_Tense--**_Tense2 (+G_NonPast +NonBefor)**
|_ Anchr: _Tense1

(13) 彼女は病気だと言った。
kanozyo-wa byooki da to itta
she -Top sick be-Pres that say-Past
'she said she was sick'
FORMULA1
|_SemHeads--言う1 (say)
|_L_Tense--_Tense1 (+Befor)
|_L_Obj--FORMULA2
|_SemHeads--病気1 (sick)
|_L_Tense--**_Tense2 (+NonBefor)**
|_ Anchr: _Tense1

In this case, both embedded tenses are anchored, since both have the anchorable feature [+NonBefor]. The English present tense is [+G_NonPast], however, so *_Tense2* denotes a time that is not before U; it is also [+NonBefor], so it also denotes a time that is not before the (past) time denoted by *_Tense1*. Consequently, the period of her sickness must overlap both the time of her saying that she was sick and the utterance time U (see also Note 8); in fact, as Enç (1987) notes, this construction has exactly that meaning. This example also illustrates the fact that a given tense may have any collection of mutually-compatible tense features, including both global and anchorable ones.

In contrast, the Japanese example (13) (the same as (1)) does not imply that the period of her sickness includes the utterance time; instead, the possibility that she is still sick at the present moment is left open, unlike (12). In our framework, this is because the Japanese present lacks a global tense feature. *_Tense2* is [+NonBefor] and not [+G_NonPast] like (12), so its only requirement is that it denote a time that is not before the time denoted by its Anchr, *_Tense1*. As indicated in the gloss, the best English translation of (13) is with the past tense. Examples like (12) and (13) illustrate precisely why the

English and Japanese present tenses are to be represented differently.

3.2. Compound tenses

One of the great insights of Reichenbach's (1947) analysis of tense is his treatment of compound tenses, such as the English present- and past-perfect. In this subsection, we outline our representation of compound tenses, which, despite notational differences, is essentially Reichenbachian.

We begin by making a formal distinction between *primary* and *secondary* tenses, the latter being tenses, such as English *have* + past participle, which require an Anchr within the same clause, the former being all others. Thus each language-particular tense must be specified as to its features, and whether it is primary or secondary. Consider the following example of the past perfect in English:

(14) He had arrived.
FORMULA1
|_SemHeads—arrive1
|_L_Tense--_Tense1 (+G_Past)
--_Tense2 (**+Befor**)
|_ **Anchr: _Tense1**

We treat English perfect constructions as consisting of two tenses: a secondary tense that is [+Befor], anchored to a primary tense, in this case simple past (hence [+G_Past]). There is no principled upper limit to the number of Tense nodes in a given clause (though particular grammars presumably impose de facto limits), though the following conditions must be met for well-formedness: (1) each clause has one and only one Tense that is not anchored within the clause (though it may be anchored outside the clause); this is the Tense that designates the "reference" time; and (2) each clause has one and only one Tense which is not the Anchr of another Tense in the same clause (though it may be the Anchr of another Tense in another clause); this is the Tense that designates the "event" time. In (14), the first condition is satisfied by *_Tense1*, and the second condition is satisfied by *_Tense2*. In the simple tense examples discussed in Section 3.1, both conditions are satisfied by the same Tense node.

The advantages of treating the perfect construction as a compound tense, instead of as a simple tense, are two-fold: (1) it allows us to distinguish English present perfect and simple past without additional features (thus helping to satisfy the design criterion (5)(a)); and (2) it captures the fact that the perfect construction co-occurs with every simple tense in English, with the same interpretation.

3.3. Survey of tenses across languages

The framework described above is not a theory of tense, in that it does not uniquely determine a representation for each grammatical tense in each language, but provides a language-neutral vocabulary for expressing differences among grammatical tenses across typologically diverse languages. To implement the framework in an NLP system, then, we need to have actual analyses of specific tenses. In this section we provide such analyses for several tenses in several languages.

3.3.1. English

The discussion above gives examples of the past, present and perfect tenses in English and their combinations. Here we give two more examples of English grammtical tenses: the future with *will*[9] and the past with *used to*.

Future: Though an argument might be made that the future with *will* is actually a compound tense, we take the simpler route here and analyze it as a distinct primary tense with the feature [+G_Future], as in the following example:

(15) You will be sick.
FORMULA1
|_SemHeads—sick1
|_L_Tense--_Tense1 (+G_Future)

Past with used to: The past tense formed with *used to*, as in *he used to work here*, like the simple past tense is [+G_Past], but differs from the simple past not only in aspectual properties (not treated here), but also in that it has the anchorable feature [+Befor]. Consider the following example:

(16) He said he used to work here.
FORMULA1
|_SemHeads—say1
|_L_Tense--_Tense1 (+G_Past)
|_L_Obj—FORMULA2
|_SemHeads—work1
|_L_Tense--_Tense2 (**+G_Past +Befor**)
|_ **Anchr: _Tense1**

Since the embedded *_Tense2* is [+Befor], it denotes a time that is not only before U, but also before the (past) time denoted by *_Tense1*. This reflects the fact that in (16), the time that he worked here must be before the time that he said he used to work here (compare to (8), above); that is, it can only mean that he said "I used to work here", and cannot mean that he said "I work here".

3.3.2. Other European languages

Apart from aspectual differences, the tense systems of Western European languages such as French, German and Spanish are very similar to that of English. The aspectual differences are of course important, and must be represented in LNS. Although a complete discussion of aspect goes beyond the scope of the present paper, we include a brief discussion of some diferences between English and Spanish here.

One notable difference between Spanish and English is that Spanish has two distinct grammatical past tenses, the perfective, or preterite, and the imperfective. The

[9] Needless to say, this is not the only way to express future-time reference in English. The simple present can sometimes be used, and there are at least two other constructions that are future only: *be going to* + infinitive, and *be about to* + infinitive. The latter construction has a different meaning from the others (immediate future), and should be distinguished, perhaps with a feature. The difference between *will* and *be going to* is hard to detect, if it exists at all, but in keeping with design criterion (5)(a) they should be distinguished in some way.

difference is entirely aspectual, and does not appear to affect the interpretation of sequence of events per se. Another notable difference between English and these other languages is that most of them use the simple grammatical present tense to refer to an event ongoing at the utterance time, as in the following Spanish example:

(17) Llueve.
rain-Pres
'It's raining.'

The simple present in English, however, cannot be used this way; English *it rains* has only a generic or habitual sense.

For both of these distinctions, a feature indicating the aspectual difference is used; in our system, the relevant features are *Discrete* and *NonDiscrete*; the former indicating that events are viewed in their entirety, the latter that events are subdivided into arbitrarily small subintervals. Thus the Spanish preterite is [+Discrete], while the imperfect is unmarked for either of these features. Also, the simple present in English is [+Discrete], while the simple present in e.g. Spanish is umarked for this feature.

Aside from such aspectual differences, the most notable tense difference between Spanish and English is that the Spanish present progressive, in contrast to the simple present, is incompatible with future time reference:

(18) Vuelvo mañana.
return-1sg tomorrow
'I return/am returning tomorrow'

(19) Estoy volviendo (*mañana).
be-1sg returning tomorrow
'I am returning tomorrow.'

This is handled by assigning the present progressive the features [+G_NonPast +NonBefor +NonAftr] (in addition to aspectual features), which differs from the simple present in being [+NonAftr]. In (19) there is no Anchr, so the [+NonAftr] feature dictates that the time referred to is not after U; i.e, is not in the future; this accounts for this tense's incompatibility with a future time adverbial.

3.3.3. Japanese

The discussion above gives some examples of the simple past and present in Japanese, analyzed in our framework as [+Befor] and [+NonBefor], respectively. Since there is no separate future tense in Japanese, future time reference is normally achieved with the simple present tense, as in the following example:

(20) 明日雨が降る。
ashita ame-ga furu
tomorrow rain-Nom fall-Pres
'Tomorrow, it will rain.'

FORMULA1
|_SemHeads—降る1 (fall)
|_L_Time—明日1 (tomorrow) (+G_Future)
|_L_Tense--_Tense1 (+NonBefor)

The feature [+NonBefor] on *_Tense1* is compatible with future time reference, as discussed in Section 3.1.3, above. The future, as opposed to present, reading of (20) comes from the presence of the adverbial *ashita* 'tomorrow'. In Section 4, we discuss how semantic tense information from adverbials is incorporated into our framework.

3.3.4. Chinese

Unlike the other languages discussed above, Chinese has no grammatical tense. As noted in the introduction vis-a-vis examples (2) and (3), semantic tense, when expressed, is often expressed via adverbials, and not with grammatical tense; this is discussed in more detail in Section 4, below. However, Chinese does have a limited number of particles, traditionally referred to as aspect markers, which, besides indicating aspect, also indicate semantic tense information. The aspectual meaning of these particles is beyond the scope of this paper, but we will discuss a few examples to show how they express semantic tense, and how that information is represented in our framework.

We discuss here the particles *le*, *guo* and *jiang*, as in the following examples:

(21) 他说他买了书
ta shuo ta mai **le** shu
he say he buy Aspect book
'He says/said that he has/had bought books.'

FORMULA1
|_SemHeads--说1 (say)
|_L_Tense--_Tense1
|_L_Obj—FORMULA2
|_SemHeads--买1 (buy)
|_L_Tense--**_Tense2 (+Befor)**
|_Anchr: _Tense1

(22) 他说他买 过书
ta shuo ta mai **guo** shu
he say he buy Aspect book
'He says/said that he has/had (once) bought books.'

FORMULA1
|_SemHeads--说1 (say)
|_L_Tense--_Tense1
|_L_Obj—FORMULA2
|_SemHeads--买1 (buy)
|_L_Tense--**_Tense2 (+Befor)**
|_Anchr: _Tense1

(23) 他说他将到美国去
ta shuo ta **jiang** dao meiguo qu
he say heAspect to US go
'He says/said that he will/would go to the US.'

FORMULA1
|_SemHeads--说1 (say)
|_L_Tense--_Tense1
|_L_Obj—FORMULA2
|_SemHeads--买1 (buy)
|_L_Tense--**_Tense2 (+Aftr)**
|_Anchr: _Tense1

In all these examples, the tense of the main clause (*_Tense1*) has no features; we take this to be the default case in Chinese, in which an unmarked clause can be interpreted as past, present or future (see the discussion of examples (2) and (3) in the Introduction, and Section 4, below). However, aspectual particles such as *le*, *guo* and *jiang* can also contribute semantic tense information, which we represent as if it were grammatical tense.

The particles *le* and *guo* are both [+Befor] (their difference is aspectual, not represented here); in (21) and (22), the embedded clause Tense is anchored to the matrix, indicating that the buying of books took place before his saying. In contrast, *jiang* in (23) is [+Aftr], so this example means that the going to the US takes place after his saying.

4. Deriving semantic tense from syntactic context

It is often the case that semantic tense information is not represented as grammatical tense per se, but can come, at least in part, from adverbials or other features of the syntactic environment. We have seen that this is one of the main sources of semantic tense information in Chinese; an example from English is *We speak tomorrow*, which is grammatically present tense (hence [+G_NonPast +NonBefor], but semantically is unambiguously about the future. To deal with this situation, we propose to augment the framework outlined in Section 3 with an additional non-tree attribute *Spcfrs*, which indicates, for a given Tense node, any other temporal expressions in the clause that contributes to the semantic tense of that clause. Like Anchr, Spcfrs is not part of the LNS tree per se, but is an annotation on the tree. The representation is given below:

(24) We speak tomorrow.
FORMULA1
|_SemHeads—speak1
|_L_Time—tomorrow1 (+G_Future)
|_L_Tense--_Tense1 (+G_NonPast +NonBefor)
                |_ Spcfrs: tomorrow1

_Tense1 has only the features of any present tense, so the representation satisfies the first design criterion (5)(a); but its Spcfrs is the adverb *tomorrow1*, which itself has the feature [+G_Future], since tomorrow is unambiguously in the future. This relation indicates to the language-independent function that derives the explicit sequential representation that the temporal reference of the clause is to a time that is after U, thus satisfying the second design criterion (5)(b).

Note that design criterion (5)(a) is satisfied in another way, as well: the structure of (24) is different from the structure of a sentence with a future tense, which presumably makes use of the feature [+G_Future] (see below); thus the distinction between the "scheduled" future (Comrie, 1985) in (24) and the more basic future of *We will speak tomorrow* is preserved.

The need for the Spcfrs relation is much more prevalent in languages that make little or no use of grammatical tense, such as Chinese. Consider the following examples:

(25) 昨天他来看我
zuotian ta lai kan wo
yesterday he come see me
'Yesterday he came to see me.'
FORMULA1
|_SemHeads--来1 (come)
|_L_Time--昨天1 (yesterday) (+G_Past)
|_L_Tense--_Tense1
                |_Spcfrs: 昨天1

(26) 明天他来看我
mingtian ta lai kan wo
tomorrow he come see me
'Tomorrow he will come to see me.'
FORMULA1
|_SemHeads--来1 (come)
|_L_Time--明天1 (tomorrow) (+G_Future)
|_L_Tense--_Tense1
                |_Spcfrs: 明天1

The Spcfrs relation thus permits specification of semantic tense features that are not expressed as grammatical tense.

5. Comparison to other frameworks

Our proposal is for a system of representation of semantic tense that is language-neutral; i.e., that represents the tense distinctions of different languages in a formal vocabulary that has the same meaning in all languages. As such, our proposal is very different from proposals to represent the semantics of tense in a particular language such as English, both in the obvious respect that we consider other languages, and in the less obvious respect that our proposal is not a semantic one in any deep sense, but rather a syntactic representation that is language-neutral, as sketched in Section 2 (Campbell & Suzuki, 2002).

As such, the nearest thing to a comparable proposal that we have encountered in the computational literature is Van Eynde (1997), which explicitly provides a Reichenbachian semantic framework for multiple languages, and incorporates information from temporal adverbs in addition to grammatical tense. Unlike our proposal, however, Van Eynde's framework is explicitly Reichenbachian, characterizing tenses in terms of three possible values for sTENSE, expressing the relation between the reference and speech times, and six values for sASPECT, expressing the relation between the event and reference times. Although our framework encodes the same essential insight, it does so without rigidly adhering to the reference time/event time distinction, which leads to a simpler representation in our view.

6. Application

Having a language-neutral representation of semantic tense has clear implications for multi-lingual applications such as MT. Consider again the Japanese example (13), in which an embedded present tense is to be translated into past tense in English. A simple transfer of the language-particular present tense yields the wrong result, since *She said she is sick* (=(12)) means something very different from (13). Instead, what needs to be transferred is the whole temporal structure of *_Tense2*, including its features and its Anchr, since this is the information that determines that it denotes a time that is before U. Such context-sensitive transfers are possible in an MT system such as that described by Richardson, *et al*. (2001).

Similarly, consider the Chinese example (25), in which there is no grammatical tense specified. A Chinese-English MT system must transfer not the grammatical tense (which yields no information whatsoever), but rather the whole temporal structure, which in this case includes its Spcfrs, in order to give the English generation system the information it needs to generate past tense.

7. Conclusion

We have presented and exemplified a framework for representing semantic tense in a language-neutral fashion, which meets the competing design criteria in (5): that each language-particular tense be reconstructible by a generation function, and that an explicit representation of temporal sequence be derivable by means of a language-independent function.

The framework we have proposed allows us to get semantic tense information from grammatical tense, or from adverbial modifiers, and represents this information in a semantically motivated, language-neutral fashion.

8. References

Campbell, R., 2002. *Language-neutral syntax*. MSR Tech Report (in preparation).

Campbell, R. & H. Suzuki. 2002. A language-neutral representation of syntactic structure. SCANALU-2002.

Comrie, B., 1985. *Tense*. Cambridge University Press.

Davidson, D. 1980. The logical form of action sentences. In D. Davidson, ed., *Essays on actions and events*, 105-122. Clarendon Press, Oxford.

Enç, M., 1987. Anchoring conditions on Tense. *Linguistic Inquiry* 18, 633-657.

Filatova, E. & E. Hovy. 2001. Assigning time-stamps to event-clauses. In Proceedings of ACL-EACL.

Heidorn, G.E. 2000. Intelligent writing assistance. In R. Dale, H. Moisl and H. Somers, eds., *Handbook of natural language processing*. Marcel Dekker, New York.

Ogihara, T. 1995. The Semantics of Tense in Embedded Clauses. *Linguistic inquiry* 26, 663-679.

Reichenbach, H. 1947. *Elements of symbolic logic*. The Free Press, New York, and Collier-Macmillan, London.

Richardson S., W. Dolan, A. Menezes and J. Pinkham. 2001. Achieving commercial-quality translation with example-based methods. In *Proceedings of the VIIIth MT summit*, Santiago de Compostela, Spain. 293-298.

Van Eynde, F. 1997. Mood, Tense & Aspect. In F. Van Eynde and P. Schmidt (eds.), *Linguistic specifications for typed feature structure frameworks*. EU Commission, Luxembourg.