A Unified Model of Tense and Modality and the Three-Tier Model of Language Use*
Naoaki Wada

1. Introduction

The aim of this article is twofold. First, I will offer a unified model of the following two theories: the compositional tense theory proposed in Wada (2001a) and developed in my subsequent studies (Wada (2001b, 2002, 2009a, 2009b, 2010a, 2010b, 2011a, 2011b, 2013, to appear)), and a theory of modality adopted to support my tense theory. Thus far, I have used the theory of modality only for supplementary purposes, especially in treating English data, so I will develop it to be motivated to provide a broader perspective. By combining the two theories, I will offer a more systematic, unifying approach to temporal and modal phenomena both in English and in Japanese with a bird's-eye view.

Second, I will demonstrate that the unified model of tense and modality to be presented is motivated and supported by (and therefore lends support to) the "three-tier model of language use" proposed by Hirose (2013). By combining the two models, we can explain, from a much broader point of view, how modality is related to situation construal and communication in English and Japanese and why modal phenomena behave differently in the two languages, especially with respect to indirect speech acts, as well as why some temporal phenomena behave the way they do in the two languages. For lack of space, the present article will merely provide a basic design for how the two models combine effectively with each other.

This article is organized as follows. Section 2 briefly outlines my tense theory. Section 3 discusses and develops the theory of modality adopted in explaining tense phenomena and its related issues within the framework of my tense theory. In section 4, I present a unified model of tense and modality on the basis of the two theories and confirm that it works. After surveying Hirose's three-tier model of language use in section 5, I show, in section 6, how the three-tier model works effectively with the unified model of tense and modality presented in section 4. Section 7 offers concluding remarks.

2. A Compositional Tense Theory
2.1. Tense-Structure Level and Tense-Interpretation Level

Let me start with our first aim. To this end, I will first briefly outline my compositional tense theory using English data. As a major characteristic, this tense theory divides the field of tense in the language system into two levels: the "tense-structure (TS)"

level, which concerns grammatical time information, and the "tense-interpretation (TI)" level, which concerns real time information. At the TS level, a given tense form represents its abstract or schematic semantic-structure, or tense structure. On the TI level (i.e. in the tense-interpretation process), the tense form is interpreted as receiving a temporal value (i.e. tense-interpretation value) under the influence of not only semantic, pragmatic, and syntactic factors but also contextual factors in the linguistic environment where it occurs. The TI level includes both of the semantic and the pragmatic interpretation of tense forms.[1] This distinction between the two tense levels enables us to deal systematically with cases where one single tense form is used to express different temporal values in different linguistic environments (we will see this in section 2.3).[2]

2.2. Absolute Tense-Component/Relative Tense-Component and Three Factors Contributing to Tense Structure

As another characteristic, the tense theory allows two tense-components constituting tense structure, i.e. the A(bsolute tense)-component and the R(elative tense)-component. The two components are defined as follows:

(1) a.   The A-component is a tense-component to which is related tense-structure information that includes a reference to the deictic center of grammatical time.[3]

  b.   The R-component is a tense-component to which is related tense-structure information that does not include a reference to the deictic center of grammatical time.

There are three types of factors contributing to tense-structure information. The first factor is a verb (predicate) stem, the core of a verb (predicate). The second factor is a tense morpheme that changes according to person, number, and mood, which is called an A(bsolute tense)-morpheme. This morpheme corresponds to what is generally called tense inflection or tense affix in English and other West European languages. The third factor is a tense morpheme that does not change according to person, number, and mood,

---

[1] Previous studies of tense and aspect have usually assumed that tense interpretation corresponds to pragmatic interpretation, and not incorporated it constructively into the core of their tense models, though there have been some exceptions, such as Discourse Representation Theory (e.g. Kamp and Reyle (1993)) and its similar frameworks (Caudal (2012) and references cited there) or a series of Declerck's tense theory (Declerck (1991, 1997, 2006)).

[2] This is a major characteristic of my tense theory because such a distinction is not made clearly in almost all tense theories (including the exceptional theories mentioned in note 1).

[3] The deictic center of grammatical time conceptually differs from the deictic center of real time, i.e. the present speaker's "now" on the real time line, though they can be and are basically identified with each other. See the discussion in section 2.3 and note 7.

which is called an R(elative tense)-morpheme. This morpheme corresponds to non-finite markers in western languages such as English and French. Non-finite markers are tense morphemes in my tense theory partly because they can express relative temporal relationships (cf. Comrie (1976:3)) and partly because treating them as tense morphemes allows us to deal with the tense systems of English and Japanese from a unified point of view (Wada (2001b, 2009a, 2011b)).

The three factors represent the following types of tense-structure information.

(2) a. A verb (predicate) stem represents an event time.
    b. An A-morpheme represents a time-sphere, i.e. a grammatical time-range, whose value is fixed in relation to the deictic center of grammatical time.
    c. An R-morpheme represents an intrinsic relationship of grammatical time between the event time and the potential time of orientation.

Only the A-morpheme is related to the A-component because, as (2b) shows, its tense-structure information includes a reference to the deictic center of grammatical time. The verb stem and the R-morpheme are related to the R-component because, as inferred from (2a, c), they do not include such a reference. Note that a verb stem must be accompanied by either an A-morpheme or an R-morpheme (at least in the languages mentioned above); a tense form consisting of an A-morpheme and a verb stem is an absolute tense form (consisting of both the A-component and the R-component), whereas a tense form consisting of an R-morpheme and a verb stem is a relative tense form (consisting only of the R-component).

In the rest of this subsection, I will explain each of (2a-c) with concrete descriptions. First, an event time is the time point or span corresponding to the relevant part (or phase) of a given situation (note that a situation covers an action, event, state of affairs, or whatever is described by a verb or verb phrase).[4] A situation can be expressed by a verb alone, a verb phrase, or a whole sentence. Take (3), for example.

(3) I fell in love with you watching "Casablanca."

The event time of the finite verb *fell* is the time length of falling in love, which corresponds to the time length of the relevant part of the situation of watching the movie "Casablanca."

Let us next consider the A-morpheme in some detail. In (3), the past tense form *fell* (an irregularly conjugated verb) is divided, in terms of tense structure, into the verb stem

---

[4] The definition of "event time" in my tense theory basically corresponds to that of "topic time" in Klein (1992, 1994) and that of the "time of the predicated situation" in Declerck (1997, 2006).

fall- and the past tense morpheme *-ed*, the latter being an A-morpheme. In present-day English, while deictic notions such as person, number, and mood are, in most cases, not represented explicitly on verbs, such a marker as the present tense morpheme *-s*, i.e. a tense morpheme integrated with the notions of "third-person," "singular," and "indicative," is existent. From this, we assume that the deictic notions are conceptually associated with all the finite verbs in (present-day) English. Under this assumption, English has two types of A-morphemes, i.e. the past tense morpheme (represented by *-ed*) and the present tense morpheme (represented by *-s*).[5] The past tense morpheme represents a past time-sphere, i.e. a grammatical time-range that is located earlier than the deictic center of grammatical time; the present tense morpheme represents a present time-sphere, i.e. a grammatical time-range that includes the deictic center of grammatical time. In the default case, the deictic center of grammatical time is identified with the deictic center of real time (we will return to this identification mechanism in the next subsection).

Let us finally consider the R-morpheme by taking the present participle *watching* in (3) as an example. At the TS level, the present participle morpheme *-ing*, which is by definition an R-morpheme because it does not change according to person, number, and mood, represents the simultaneity in grammatical time between the event time and the potential time of orientation, i.e. an evaluation time whose specific value will be determined in the tense-interpretation process (on the TI level) depending on the characteristics of the relevant linguistic environment (for the tense-interpretation process of this non-finite form, see note 12).

2.3. Tense-Interpretation Process

Next, we will see how the temporal value of a given tense form is identified in the tense-interpretation process. Consider, first, the English past tense form as an example of the absolute tense form. The tense structure of the English past tense form is such that the event time represented by the verb stem is located somewhere in the past time-sphere represented by the past tense morpheme *-ed*, as schematized in Figure 1.
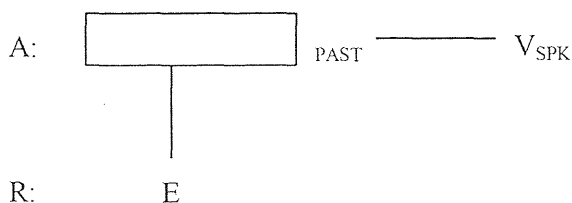
A: ⎡⎯⎯⎯⎯⎯⎯⎯⎤ PAST ⎯⎯⎯⎯⎯ V_{SPK}

R: E

Fig.1: Tense Structure of the English Past Tense Form

---

[5] In this tense theory, *will*, often considered to be a future tense marker, is not an A-morpheme. The inflectional ending of the simple future in French is an A-morpheme (e.g. *-rai* in *Je viendrai* 'I will come'). For details, see Wada (2013).

A and R stand for the A-component and the R-component, respectively. V_{SPK} symbolizes the speaker's t(emporal)-viewpoint, the rectangle with subscript PAST denotes a past time-sphere, and E indicates the event time. The horizontal and vertical lines, respectively, represent a before/after relationship and a simultaneous one (including one of inclusion). This tense structure is meant to be shared by all instances of the English past tense form and motivates us to use this form.

In the default case, the speaker's t-viewpoint (V_{SPK}), i.e. the deictic center of grammatical time, fuses with his/her consciousness (C_{SPK}), i.e. part of the brain engaged in any type of cognitive activity, such as uttering or thinking, which is by definition always existent at speech time (S), i.e. the deictic center of real time.[6] Because English finite forms include the speaker's t-viewpoint in their tense structure, the fusion under consideration normally (i.e. in the default case) occurs in the past tense form, which is chosen with speech time serving as the base point for the tense-form choice, irrespective of whether the tense form occurs in main or subordinate clauses.

To illustrate the point, consider (4):

(4)   John played baseball.

The temporal structure (i.e. the semantic structure of a tense form on the TI level functioning as the "template" for calculating its temporal value) of the past tense form *played* is, in the default case, schematized in Figure 2.

PAST TIME-AREA | $S(C_{SPK})$

A: V_{SPK}
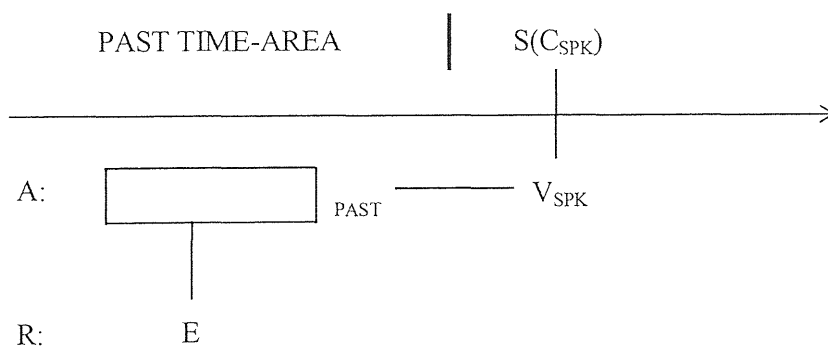   PAST

R: E

Fig.2: Temporal Structure of the English Past Tense Form (Default Case)

The bold vertical line divides two time-areas (real time-ranges). Here, the speaker's t-viewpoint in the tense structure of the past tense form *played* fuses with his/her

---

[6] In my theory, speech time is used in a broader sense, including not only the time of the speaking but also the time of the thinking.

consciousness, so it is situated at speech time;[7] as a result, the past time-sphere (a grammatical time-range) corresponds to the past time-area (a real time-range), and thus, the event time occurs in the past on the real time line. In a case containing a finite tense form in a main clause, e.g. (4), the time of orientation, i.e. the base time to evaluate (the position of) the event time, is normally identified with speech time. The length of the event time of *played* itself is vague. However, if a time-duration adverbial like *for two hours* is added, the time length is identified; if a time-specifying adverbial like *yesterday* is added, the time length is maximally one day long and the time position is specified (cf. Rathert (2012)).

Let us next take a brief look at a non-default case of the past tense form, where the speaker's t-viewpoint is situated at a time other than speech time. Consider (5):

(5) John will say on the twentieth of May that he arrived.

(adapted from Comrie (1985:112))

The temporal structure of the past tense form *arrived* is schematized in Figure 3.
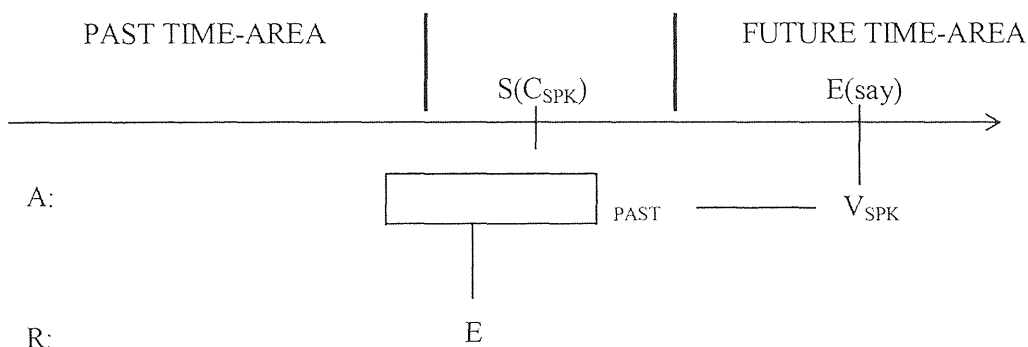

Fig.3: Temporal Structure of the English Past Tense Form (Non-Default Case)

In this case, the speaker's t-viewpoint in the tense structure of the past tense form fuses with the consciousness of the original speaker *John* at the expected time of his utterance (represented by *say* in (5)).[8] Therefore, the past time-sphere does not necessarily

---

[7] In English, non-default or marked cases are only cases such as the historical present (see note 10) or the indirect-speech complement of a reporting clause referring to the future (see (5)); almost all the cases constitute the default case. Therefore, basically, the speaker's t-viewpoint fuses with his/her consciousness and is situated at speech time.

[8] A major reason why such a fusion occurs in the indirect-speech complement of a verb of saying referring to the future is that if the reporter, on his/her own, judges the truth value of, and "rearranges," what has not been yet but will be uttered by the expected original speaker, it is highly possible that he or she will skew the originally intended content (cf. Harder (1996)).

correspond to the past time-area, but can cover any time-area. This is inferred from the fact that there is no direct relation between the past time-sphere and speech time in the figure. What deserves attention here is that this past tense form shares the same tense structure with the one in (4).

We will then move on to a consideration of a combined tense form consisting of a finite and a non-finite verb. Take (6), a present-perfect sentence, as an example.

(6) Adolph has played the saxophone.

In treating an example like this, we must note that my tense theory adopts a hypothesis as to auxiliaries like (7):

(7) Auxiliary as well as lexical verbs can have their own event times.[9]

Under this hypothesis, sentence (6) contains two event times, i.e. the event time of the perfect auxiliary *has* and the event time of the past participle *played*.

With this in mind, let us present the temporal structure of the present perfect form *has played* in the default case, which is schematized in Figure 4.

PAST TIME-AREA $S(C_{SPK})$

A: V_{SPK} PRES
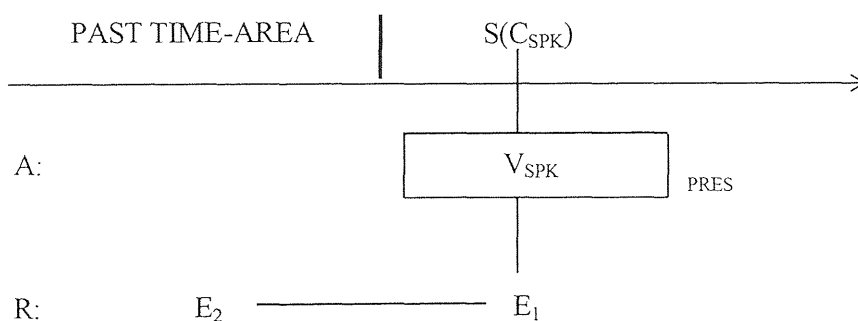
R: E_2 —————— E_1

Fig.4: Temporal Structure of the Present Perfect Form (Default Case)

I will start with the interpretation mechanism of the perfect *have* in the present tense (an absolute tense form). The tense structure of the English present tense form *has* is such that the event time represented by the verb stem is located somewhere in the present time-sphere represented by the present tense morpheme -*s*. Here again, in the tense-interpretation process the speaker's t-viewpoint fuses with his/her consciousness existent

---

[9] This hypothesis is also assumed in Janssen (1994, 1996) and Nakau (1994), to which I owe much. For arguments to support this hypothesis in terms of prototype theory, see Wada (2001a:Ch. 2).

at speech time by default;[10] the present time-sphere (a grammatical time-range) in this tense form covers the present time-area and the future time-area, i.e. real time-ranges. To make the discussion simpler, I simply assume here that the event time of the present tense form in English independent clauses basically obtains at or includes speech time and this is the case with the present tense form in question.[11]

I will next consider the past participle *played*, a non-finite verb. In my theory, the past participle morpheme *-en* (an R-morpheme) represents an anterior relationship in grammatical time between the event time and the potential time of orientation at the TS level. In interpreting this tense form (i.e. on the TI level), the potential time of orientation is identified with the time of the finite (matrix) verb *has* as the head of the verb phrase because of the characteristic of this linguistic environment, i.e. the complement position of the matrix verb. Therefore, the event time of the perfect *have* serves as the time of orientation for evaluating the event time of the past participle *played*.[12] Because the event time of the perfect *have* (E_1) is simultaneous with speech time (S), the event time of the past participle *played* (E_2) is interpreted as obtaining in the past time-area. The fact that E_1 shares the same time with speech time indicates that the present perfect form has the so-called current relevance, which the simple past form does not have (see Figure 2 above).[13]

3. A Theory of Modality

Next, I will develop the theory of modality adopted in my previous studies so that it can provide a broader perspective and be linked with the tense theory in a more motivated

---

[10] An example of a non-default case where the present tense form is chosen with a time other than speech time serving as the base point for the tense-form choice is the historical present. In this case, the speaker's t-viewpoint is situated at a time in the past time-area. See especially Wada (2011a:45) for details.

[11] Because the present time-sphere covers both the present and the future time-area in the default case, the event time of the present tense form can theoretically obtain either at speech time or in the future. Therefore, the event time can theoretically hold in the future. However, the event time of a present tense form in an independent clause is normally interpreted as obtaining at or including speech time in English. This is partly because, as we will see in sections 3 and 4, assertive modality, or assertion, is associated with the finite form in interpreting the sentence. In addition, because of the idiosyncratic features of the perfect *have* (cf. Huddleston (1977)), the relevant event time in (6) must be simultaneous with speech time.

[12] In a similar way, the event time of the present participle *watching* in (3) in the main text is interpreted as simultaneous with the event time of the finite (main) verb *fell* as the time of orientation because a participial construction is usually syntactically and semantically "subordinated" to the main clause.

[13] In the case of the continuative use, as in *I have known her since 1992*, E_2 is interpreted as reaching E_1 in the course of tense interpretation.

way.[14] (Here again, I will use English data to illustrate the point.) The modality theory presupposes the following hypothesis:

(8) A sentential utterance is semantically divided into the speaker's attitude domain and the proposition domain.

This is a hypothesis about the semantic decomposition of a sentential utterance (including one in the speaker's mind). The speaker's attitude (SA) domain is concerned with the subjective aspect of an utterance (the term "subjective" is used in the sense of 'on the part of, or belonging to, the speaker as the subject of cognitive activities'), whereas the proposition (P) domain is concerned with the objective aspect of it (the term "objective" is used in the sense of 'detached from the speaker'). Linguistic elements belonging to the SA domain reflect the speaker's mental attitude or state holding at the time of utterance or thought (normally identified as speech time), and those belonging to the P domain are situations or scenes to be described, i.e. objective content.[15]

Hypothesis (8) appears to be similar to one adopted in Searlean speech act theory, where the utterance of a sentence must include both an illocutionary force and propositional content (cf. Searle (1969, 1979)). However, the SA domain includes not only illocutionary forces (or points) but also what Verstraete (2001) calls "modal performativity," i.e. the performativity involved in subjective modality that expresses the speaker's particular position of commitment with respect to the propositional content of the utterance (Verstraete (2001:1517)). In this way, hypothesis (8) is motivated to have a broader perspective. On this basis, three major types of modality (i.e. epistemic, deontic, and dynamic modality) are partitioned in such a way that besides speech acts, epistemic and deontic modality as well as evidentiality are sorted into the SA domain, whereas dynamic modality as well as propositional content are sorted into the P domain.[16,17] Our claim that not only speech acts but also epistemic modality, deontic modality, and evidentiality are subsumed under the notion of "speaker's attitude" is in keeping with

---

[14] Thus far, many scholars have developed their own theories of modality from various points of view. Verstraete (2001) contains a good summary and comparison of some major approaches to modality.

[15] In main or independent clauses, the time of utterance or thought in question is speech time. As we will see later, however, in the case of indirect speech complements, for example, elements belonging to the SA domain (such as epistemic or deontic modality) can reflect the original speaker's mental attitude at the time of the original utterance or thought.

[16] For what elements are included in the notion of evidentiality, see Watanabe (2004), Nuyts (2005), and de Haan (2012).

[17] This partition is slightly different from that of Palmer (2001), who considers both epistemic modality and evidentiality (what he calls "evidential modality") to form a group, i.e. "propositional modality," and distinguishes deontic modality from the two types of modality to group the former and dynamic modality into "event modality."

Nuyts's (2005:21-23) observation that the latter three form a group and "involve (different types of) speaker attitudes towards the state of affairs."[18] What is to be stressed here is that in this theory, epistemic and deontic modality are defined as (different types of) the speaker's mental attitudes, including those toward the situation (such as Verstraete's (2001) modal performativity) and those toward the addressee (related, at least in part, to illocutionary forces). Note that epistemic modality is an indication of the degree of probability of the situation, whereas deontic modality is an indication of the degree of moral desirability of the situation (Nuyts (2005)). This way of division of modality is the developed version of the theory of modality.

To illustrate the point, let us first consider (9):

(9) a. John may come tomorrow.
    b. Mary must be at her office now.

An utterance of sentence (9a) consists of the propositional content, i.e. the situation of John's coming tomorrow, and the speaker's mental attitude toward the situation, i.e. modality of possibility (a type of epistemic modality). Similarly, an utterance of sentence (9b) is composed of the situation of Mary's being at her office now (i.e. the propositional content) and the modality of logical necessity (i.e. the speaker's mental attitude).

I will next consider sentences containing *will* (*will*-sentences), which have been treated differently in the literature. Observe (10):

(10) a. Toru will go to Germany.
     b. Yoko will be at home now.

In my tense theory, *will* is not a tense marker, but a modal. The *will* in (10b) expresses epistemic modality, used to infer the present situation. This position of mine is the same as that of almost all studies on modality. By contrast, the *will* in (10a) is often treated as a future tense marker in the literature, but in my theory of modality (working in a close relationship with my tense theory) this *will* is also a modal expressing predictive modality (prediction). This is because prediction is defined as a mental attitude in which the speaker forecasts on a reasoned basis—"forecast" means "calculate or estimate something conjecturally" (Wada (2011a:40); cf. also Close (1977:131))—and can be treated as a type

---

[18] Nuyts (2005:18) considers the three types of modality (i.e. epistemic, deontic, and dynamic modality) as well as other elements such as evidentiality and expressions of time or space to be "qualificational" categories, i.e. categories "which speakers can express with respect to the states of affairs they are talking about." This system of qualificational categories lays a necessary foundation for his claim in the main text and leads him to the decomposition of the traditional notion of modality.

of epistemic modality sorted into the SA domain.[19] In my theory of modality, both of the *will*'s with present and future time reference are grouped together to express the same type of modality, i.e. prediction (cf. Leech (2004)); this unified treatment is a merit of the theory.

I now move to a discussion of sentences containing dynamic modals, as exemplified in (11):

(11) a.   Hanako can play the harp very well.

b.   I will go to Europe next month.

In (11a), not only the propositional content but also the modality of ability expressed by the dynamic modal *can* are sorted into the P domain. A question, then, arises as to what is sorted into the SA domain to meet hypothesis (8). To answer this question, the present theory assumes that assertive modality (assertion), a type of epistemic modality, is conveyed in interpreting this type of sentence.[20] Assertion is defined as a mental attitude in which the speaker (subjectively) states the propositional content as a fact (Wada (2011a:39); cf. also Searle (1969:29)), and its nature matches up with the nature of "subjective modality" in the sense of Verstraete (2001) mentioned above. In English, assertion is a non-linguistic element belonging to the SA domain, but the existence of this notion in sentences of the type under discussion is a natural consequence of our position that epistemic modality is a speaker's mental attitude as well as our hypothesis that every sentential utterance involves a speaker's attitude. From these observations, sentence (11a) is interpreted in such a way that the speaker asserts that the present situation of Hanako's having the ability to play the harp very well is true.

By the same token, in (11b), both the propositional content and the subject's volition (a type of dynamic modality) belong to the P domain, so that assertive modality is interpreted as accompanying the utterance. The speaker asserts that the present situation of the subject's having the volition to go to Europe next month is true.

The discussion thus far leads us to claim that in the theory of modality sentences without explicit modal elements, i.e. unmodalized sentences, can convey assertive modality. This type of modality is unmarked because normally the speaker has no doubt

---

[19] My claim that prediction is a type of epistemic modality is also supported by Nuyts's (2005:23) statement that "epistemic modality involves an explicitation of the degree of 'existential' commitment of the speaker to the state of affairs, i.e. the extent to which (s)he believes the state of affairs has been or will be realised in the 'real world'" (my emphasis). For further arguments for the position that future *will* expresses predictive modality, see Wada (2011a, to appear) and the references cited therein.

[20] Langacker (2008:472) uses "statement" instead of "assertion" because he considers the latter to be "overly pretentious." See also Palmer (2001:64).

about the truth of the propositional content that he or she is describing or construing and states it straightforwardly. Consider (12):[21]

(12) a.　Leon is a hit-man.
　　 b.　Leon may be a hit-man.

Both (12a) and (12b) have the same propositional content, i.e. the situation of Leon's being a hit-man. With respect to the SA domain, it is occupied by assertive modality in (12a) and by modality of possibility in (12b).

4. A Unified Model of Tense and Modality

Having seen the tense theory and the modality theory to be combined with each other, I will now present a unified model of tense and modality and show how the model works in explaining temporal and modal phenomena in English and Japanese more systematically from a unified point of view. The models for English and Japanese are diagrammed in Figure 5(i) and Figure 5(ii), respectively:

(i) English　　　　　　　　　(ii) Japanese

$$\boxed{\text{SPK}} \rightarrow [_{SA}\ \alpha_i\ [_P\ \beta_{i/j}\ ]\ ] \qquad \boxed{\text{SPK}} \rightarrow [_{SA}\ \alpha_i\ [_P\ \beta_{i/j}\ ]\ ]$$

English: $C\ \&\ V^t$

Japanese: $C\ \&\ V$

A: $\Box$　X

English R: $^{f}E_i{}^{0<n}\ \&\ ^{nf}E_j{}^{0<n}$

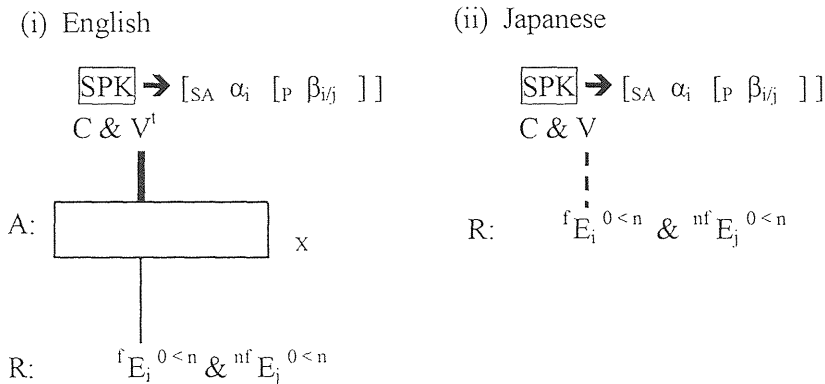Japanese R: $^{f}E_i{}^{0<n}\ \&\ ^{nf}E_j{}^{0<n}$

Fig.5: A Unified Model of Tense and Modality (i) English (ii) Japanese

Let us start with common denominators between the two figures. The surrounded SPK symbolizes the speaker; the C and V (including V^t) beneath it represent the speaker's consciousness and viewpoint, respectively; the bold arrow indicates the speaker's

---

[21] In such a case as (i),

(i) Maybe, Leon is a hit-man.

the SA domain is occupied by the modality of possibility expressed by the modal adverb *maybe*. Assertion is inserted into the SA domain only when there is no explicit indication of modal expressions in the sentence.

involvement in the utterance; the subscript SA and P suggest the speaker's attitude (SA) domain (whose elements are represented by α) and the proposition (P) domain (whose elements are represented by β), respectively; ^fE and ^{nf}E denote the event time of a finite form and that of a non-finite form, respectively; as to the right side of the upper part of E, the number on the left side of the inequality sign shows the minimum number of the event time involved, while the number on the right side of the inequality sign shows the maximum number (n indicates any number).[22] The ampersand between ^fE and ^{nf}E merely suggests that the verbs (predicates) coexist with each other, irrespective of whether they are in the same verb phrase or projected into different syntactic positions. The indices i and j are "correspondence" markers: elements with an index in the SA and the P domains (e.g. α_i and β_j) are linked to those with the same index in the R-component of tense structure (e.g. E_i and E_j). To be more specific, ^fE can be either the event time of a verb belonging to the SA domain or that of a verb belonging to the P domain, whereas ^{nf}E is the event time of a verb belonging to the P domain.[23]

Let us next consider the non-common elements. In the English case (i), V^t (also shown as V_{SPK}) symbolizes the speaker's t-viewpoint and the rectangle with subscript X denotes a time-sphere, and the bold line between them indicates that the speaker's t-viewpoint is necessarily involved in the use of an absolute tense form (a tense form with a time-sphere); the fact that ^fE is linked by a solid line to the rectangle implies that the event time obtains somewhere in the time-sphere.

In the Japanese case (ii), V symbolizes a viewpoint of the speaker, which functions as the base point from which to evaluate or see the target object or situation;[24] this type of speaker's viewpoint can also serve as the base point to choose a tense form and is put on a point on the time line depending on the characteristics of the linguistic environment in which the tense form occurs. The broken line between V and the R-component means that either type of event time in the R-component is linked to this type of speaker's

---

[22] In most cases, a given sentence must include at least one finite form, but some special sentences, i.e. idiomatic expressions, may contain only a non-finite verb, as in (i):

    (i)   a.   How about going to the movies tonight?   (Searle (1979:40))
           b.   Why not stop here?   (Searle (1979:52))

[23] Strictly speaking, an expression like *strictly speaking*, for instance, might show that the event time of a non-finite verb is linked to the SA domain because such an expression seems to express a speaker's mental attitude. For this, I tentatively consider such expressions to be idiomatic and not to include "pure" non-finite verbs. I thank Junya Watanabe (personal communication) for bringing this to my attention.

[24] The type of speaker's viewpoint under consideration here is equivalent to the speaker's viewpoint of situation description (SD-viewpoint) in my other studies. The time on which is put the speaker's SD-viewpoint serves as the base point in time to choose a tense form as well as the time of orientation for evaluating (the position of) the event time in Japanese. In English, it serves only as the time of orientation for evaluating the event time of both finite and non-finite verbs. See especially Wada (2009a) for the detailed interpretation mechanisms with this notion.

viewpoint when evaluated. The point here is that Japanese does not have absolute tense forms; even finite forms are relative tense forms (tense forms without a time-sphere).

For a better understanding of the model, let us consider concrete examples. I will begin by considering how English examples are treated in terms of Figure 5(i). The examples to be considered are shown in (13):

(13) a.　John played baseball.　　　　　　　　　　　　　　　　(= (4))

　　　b.　Toru will go to Germany.　　　　　　　　　　　　　　(= (10a))

　　　c.　If the weather is fine tomorrow, they will go on a picnic.

Sentence (13a) has only one verb and thus one event time; the verb is in the past tense and an absolute tense form (i.e. an English finite form). The variable on the time-sphere (symbolized by X) is fixed to PAST; the number of the event time of finite forms (^{f}E) is one, and that of the event time of non-finite forms (^{nf}E) is zero. On the TI level, the past time-sphere (a grammatical time-range) corresponds to the past time-area (a real time-range) because of the fusion of the speaker's t-viewpoint (V^{t}) and his/her consciousness (C) existent at speech time, and therefore, the event time obtains in the past on the real time line. Because the event time involved is that of the lexical verb *play* (constituting part of the propositional content), it is linked to the P-domain element β, i.e. the playing, because of the correspondence index i. In interpreting the uttered sentence, assertion (the unmarked mental attitude of the speaker toward the situation to be described) is inserted into the SA-domain as α because there is no explicit modal expression.[25]

Let us turn to sentence (13b), which contains two verbs. The modal *will* is finite, having its own event time, i.e. ^{f}E, as well as a time-sphere in its tense structure;[26] the bare infinitive *go* is non-finite, having the event time (symbolized by ^{nf}E) and the potential time of orientation in its tense structure. Because the finite verb *will* is present, the variable on the time-sphere is fixed to PRES(ENT); when the *will* is interpreted as expressing the speaker's prediction, i.e. a speaker's mental attitude at speech time, its event time is linked

---

[25] From the point of view of the speaker, he or she utters or conveys the sentence in question assuming that the addressee will interpret it in the way pointed out in the main text. The same is true for the other examples in this article.

[26] A piece of evidence for the view that *will* (as well as other modals) is a finite verb is the fact that in English independent clauses finite verbs must occupy the left-most position of a verb phrase, as exemplified in (i):

(i)　　John tried to turn off the light.

Because *will* in (13b) occupies the left-most position of the verb phrase of a (non-idiomatic) independent clause, it should be regarded as a finite verb.

to the SA domain element α (i.e. prediction), whereas the event time of the lexical verb *go* is linked to the P-domain element β (i.e. the going). When the *will* is interpreted as expressing the subject's volition, i.e. a subject's present state of mind (Leech (2004:62)), both event times are linked to the P-domain element β, and assertion is inserted into the SA-domain in the interpretation process. In either reading, on the TI level the potential time of orientation associated with the non-finite verb *go* is identified with the event time of the finite verb *will* because of the complement position of the modal as matrix verb; the event time of *go* (a non-stative verb) is posterior to that of *will*, which in turn is simultaneous with speech time because of the characteristics of the modality involved (prediction or volition). The posterior relationship is due to Duffley's (1992) view that with the bare infinitive following a modal, a non-stative verb represents posteriority, but a stative verb represents either simultaneity or posteriority, relative to the time of the modal.

Let us finally consider (13c), which consists of two clauses, i.e. an *if*-clause (a conditional clause) and the main clause. The *if*-clause in (13c) is what I call a "semantically deficient" clause (Wada (2011b)), a clause consisting only of the P domain, and expresses a (direct) cause-effect relationship between the protasis and the apodosis. In sentences with this type of *if*-clause, the propositional content of the *if*-clause (i.e. the protasis) is incorporated into (and thus forms part of) the propositional content of the main clause (i.e. the apodosis), and therefore, the composite proposition is under the "scope" of an SA-domain element of the main clause (for arguments for this, see Allen (1966), Haegeman and Wekker (1984), and Wada (2011b) as well as the discussion in section 6.1.1 below). Given that one sentential utterance consists of both SA-domain and P-domain elements (see hypothesis (8)), (13c) forms one utterance composed of the prediction (a type of epistemic modality) represented by *will* in the main clause (belonging to the SA-domain) and the composite proposition in which the weather's being fine tomorrow causes their going on a picnic (belonging to the P domain). Because the two finite verbs (i.e. *will* and *is*) are both in the present tense, the variable on the time-sphere is fixed to PRES in both cases. Whereas *will* in the main clause expresses predictive modality (whose event time is linked to the SA-domain element α), *is* in the *if*-clause is not accompanied by assertive modality because the *if*-clause does not have its own SA domain (cf. Searle (1969:29)). Both of the event times of the lexical verbs *is* and *go* are linked to the P-domain element β. The event time of *will* holds at speech time for the reason mentioned above; the event time of the non-stative verb *go* in the bare infinitive is posterior to it (for the reason stated above) and therefore obtains in the future time-area. The event time of *is* in the subordinate clause requires the event time of *go* in the main clause to be the time of orientation because they are both elements constituting the composite proposition representing a causal (and thus close) relationship: because of the causal relationship between the two situations (i.e. the weather's being fine and their

going on a picnic), the event time of *is* comes just before or is in a "sloppy simultaneous" relationship to the event time of *go*.[27]

Now, let us move on to consider how Japanese examples, as in (14), are treated in terms of Figure 5(ii).[28]

(14)　a.　Kinoo　　　Akane-wa　　Bahha-no　　furuutokyoosookyoku-o　　hii-ta.
　　　　　　yesterday　Akane-TOP　Bach-GEN　flute concerto-ACC　　　　play-ANT
　　　　　　'Akane played Bach's flute concerto yesterday.'

　　　　b.　Asu　　　　Hanako-wa　　Henderu-no　　haapukyoosookyoku-o
　　　　　　tomorrow　Hanako-TOP　Handel-GEN　harp concerto-ACC
　　　　　　hiku　　daroo.
　　　　　　play　　will
　　　　　　'Hanako will play Handel's harp concerto tomorrow.'

　　　　c.　Asu　　　　hare-ta-ra　　　karera-wa　　pikunikku-ni　iku　　daroo.
　　　　　　tomorrow　be fine-ANT-if　they-TOP　　picnic-to　　　go　　will
　　　　　　'If the weather is fine tomorrow, they will go on a picnic.'

　　　　d.　Kaesaru-wa　Garia-o　　　seifukusi-ta.　　Kare-wa　Buritania　mo
　　　　　　Caesar-TOP　Galia-ACC　conquer-ANT　he-TOP　　Britannia　too
　　　　　　seifukusuru　daroo.
　　　　　　conquer　　　will
　　　　　　'Caesar conquered Galia. He would conquer Britannia, too.'

Sentence (14a) has only one verb and thus one event time. The verb *hiita* 'played' is finite, but not a past tense form (an absolute tense form); it is a relative tense form. This is because *-ta* is an R-morpheme in that it does not change according to person, number, and mood (see Wada (2001b, 2009a, 2011b)). This tense morpheme, at the TS level, represents an intrinsic relationship of anteriority in grammatical time between the event time and the potential time of orientation (this is why the gloss of *-ta* in Japanese examples is ANT, not PAST). Because Japanese finite forms are tense forms with an R-morpheme and thus relative tense forms, they do not include the speaker's t-viewpoint (V_{SPK} or V^t) in their tense structures; the potential time of orientation evoked by the R-morpheme *-ta* is, on the TI level, identified with speech time by virtue of the characteristics of independent clauses in the conversational mode (we are assuming in this article that unless otherwise noted, examples are from the conversational or pseudo-

---

[27] The term "sloppy simultaneous" is due to Declerck (1991, 1997, 2006).

[28] The following abbreviations are used in the glosses of Japanese examples in this article: ACC = accusative, ANT = anteriority, CF = continued form, COMP = complementizer, COP = copula, GEN = genitive, NOM = nominative, NON-A = non-anteriority, Q = question marker, QUOT = quotative, SFP = sentence-final particle, STAT = stative, TOP = topic.

conversational mode). Therefore, the event time of the finite verb *hiita* is located in the past time-area. This is why a finite verb ending in *-ta* (i.e. *-ta* form) has been treated as a past tense form in many previous studies, but it should not be in terms of my model. In this case, the speaker's viewpoint (symbolized by V) happens to be at speech time, which serves as the base point in time for evaluating the event time associated with the situation of playing a flute concerto, because of the linguistic environment in which the *-ta* form occurs. This holds true of the so-called Japanese present tense form (i.e. *-ru* form), another Japanese finite form (see Wada (2001b, 2009a, 2011b) for further discussion).

This claim is verified by the fact that in sentence (14d), which is presented as an example of the narrative mode, the *-ru* form *seifukusuru* 'conquer' (which represents an intrinsic relationship of non-anteriority in grammatical time between the event time and the potential time of orientation at the TS level) is interpreted in such a way that its event time is posterior to the event time of the preceding finite verb *seifukusita* 'conquered,' not to speech time (i.e. the time of the narration). This relationship of posteriority is due to the general view that the event time of the *-ru* form of a non-stative predicate (e.g. *seifukusuru* in (14d)) is interpreted as posterior to the time of orientation, whereas the event time of the *-ru* form of a stative predicate (e.g. *sitteiru* 'know,' *sukida* 'like') is usually interpreted as simultaneous with the time of orientation, although the posterior relationship is also possible with many stative predicates. The speaker's viewpoint for evaluating the situation of the second sentence is put on the event time of the *-ta* form *seifukusita* because of the characteristics of this linguistic environment. This is possible because Japanese finite forms are relative tense forms and do not include any time-sphere in their tense structures; since the *-ru* form does not include in its tense structure the present time-sphere corresponding normally to the non-past time-area, it can be chosen even in the case of the past time reference. By contrast, since English finite forms are absolute tense forms, they include the speaker's t-viewpoint, which will on the TI level fuse with his/her consciousness at speech time in the default case, basically irrespective of the type of linguistic environment. Thus, in (14d), i.e. a past context, as shown in the English translation, the past tense form *would* (*conquer*) is chosen because it includes in its tense structure the past time-sphere, i.e. a time-sphere prior to the speaker's t-viewpoint that will be situated at the time of the narration in the tense-interpretation process. It should also be noted that in both of sentence (14a) and the first sentence of (14d), the event time of the finite verb is linked to the P-domain element β, and assertion (i.e. assertive modality) is inserted into the SA domain as α in the course of interpretation.

We turn now to (14b). As with the English example in (13b) above, the modal verb *daroo* 'will' expresses prediction, which by definition obtains at speech time, and thus its event time is simultaneous with speech time; the event time of the lexical verb *hiku* 'play' (a non-stative verb) is interpreted as posterior to the event time of *daroo* as the time of

orientation and therefore located in the future time-area.[29] The event time associated with *daroo* is linked to the SA-domain element α (i.e. prediction), whereas the event time of *hiku* is linked to the P-domain element β (i.e. the playing).

Finally, we will consider (14c). As with the English example in (13c) above, sentence (14c) also consists of two clauses, i.e. a -*(ta)ra*-clause (a conditional clause) and the main clause, but forms one utterance. I argued in Wada (2011b) that Japanese -*(ta)ra* clauses are semantically deficient clauses (see also the related discussion in section 6.1.2 below).[30] Therefore, the propositional content of the -*(ta)ra* clause is incorporated into (and thus forms part of) the propositional content of the main clause, and the composite proposition is under the "scope" of the predictive modality represented by the modal *daroo* 'will.' The event time of *daroo* is linked to the SA-domain element α (i.e. prediction); by contrast, the event time of *hare(ta)* 'be fine' in the conditional clause and that of *iku* 'go' in the main clause are both linked to the P-domain element β (i.e. the being fine causing the going). The event time of *daroo* is seen as simultaneous with speech time because of the nature of epistemic modality (in the main clause) and the event time of the situation described by the lexical verb *iku* (a non-stative verb) is interpreted as posterior to the event time of *daroo* as the time of orientation. The event time of *hare(ta)* comes just before or is in a sloppy-simultaneous relationship to the event time of *iku* because of the causal relation between the two situations, i.e. the weather's being fine and their going on a picnic. In this way, the two situations are interpreted as occurring in the future.

In this section, we have seen how our unified model of tense and modality works. Recall here that the model, especially the field of modality, has been developed in

---

[29] Here, I assume that the lexical verb before the modal *daroo* is finite. A reason for this assumption is that the so-called -*ru*/-*ta* alternation is possible with the lexical verb followed by the modal, as in *hik-u/hii-ta daroo* 'will {play/have played},' which I think suggests that the lexical verb serves as a finite verb in the complement clause (Yukio Hirose (personal communication)). I leave to future research what the syntactic structure of this combination is like. I also assume that the lexical verb in this environment is not accompanied by any speaker's attitude, because basically there is only one mental attitude for one speaker in one utterance and the situation in this environment is semantically incorporated into the scope of the SA domain element represented by the modal *daroo* in that this modal cannot be used alone but requires a verbal complement over which it has an influence and thus they constitute a "unit." (Note, in passing, that two modals can be used consecutively in a single verb phrase, as in *Yooko-ga kuru kamosirenai daroo* 'It will be possible for Yoko to come'; here, the modal *kamosirenai* 'may' is assumed to express dynamic modality and therefore no speaker's attitude accompanies this modal, which is consistent with my assumption here. The claim here is supported by the fact that the example in question can be paraphrased as *Yooko-ga kuru kanoosei-ga aru daroo* 'There will be a possibility for Yoko to come,' where *kanoosei-ga aru* 'be a possibility' (a lexical predicate) is a paraphrase of the modal *kamosirenai*.) Although we need to offer a syntactic background for this assumption, Japanese has other cases like this, as in *Kanozyo-ga saru no-o mita* '(I) saw her leave.' The finite form *saru* 'leave' is not accompanied by any speaker's attitude because the complement situation is only a reflection of the speaker's visual scene. In this connection, see the discussion about example (47a) in the main text.

[30] See also Masuoka (1991, 1997) for further arguments for this position.

relationship to the semantic aspect of sentential utterances. This implies that the model has an affinity for the three-tier model of language use proposed by Hirose (2013) because the latter is a model of situation construal and communication, i.e. notions closely related to sentential utterances. In the following two sections, I will combine the two models and show how the combined model works, which is the second aim of this article.

5. The Three-Tier Model of Language Use

I will first outline Hirose's (2013) three-tier model of language use briefly in this section, and then show, in section 6, that it motivates and lends support to (and is supported by) the unified model of tense and modality proposed above.

5.1. Public Self/Private Self and Public-Self Centered Language/Private-Self Centered Language

Let us first observe two key concepts for Hirose's three-tier model, i.e. public self and private self, which are two aspects of the speaker (Hirose (1995, 1997, 2000, 2002); Hasegawa and Hirose (2005); Hirose and Hasegawa (2010)). The public self is the subject of (linguistic) communication and has an addressee in mind; the private self is the subject of thinking or mental representation and has no addressee in mind when engaged in linguistic acts. He has developed a model of language typology that can deal with linguistic differences in terms of which of the two aspects of the speaker the system of a given language is centered around. In particular, he argues that English is a public-self centered language and Japanese is a private-self centered language. As to arguments for this distinction, I consider just two of them, which are directly relevant for our purposes (I would like to refer readers to his studies mentioned above for other arguments).[31]

The first argument is concerned with whether a language has an established system of grammatical person as reflected in the distinction between first, second, and third person. In languages with this deictic system, speaker and addressee are first and second persons, grouped together as the direct participants in a speech act (cf. Benveniste (1971)), which implies that the speaker assumes the presence of an addressee in uttering (cf. also Langacker's (2008) "grounding" model). In this sense, speaker and addressee are on equal terms in these languages. This enables us to claim that English "features" the public self in its language system and is therefore a public-self centered language. By contrast,

---

[31] The arguments provided by Hirose—at least some—seem to show that other West European languages such as French, German, Dutch, and Spanish are also public-self centered languages. Nevertheless, there are differences in tense and mood phenomena among these languages. I have tried to explain these differences in terms of the difference of degree of public-self centeredness and shown that in so doing, the notion of "C-gravitation," i.e. the "gravitation" of linguistic forms or their semantic range toward the consciousness of the speaker as public self, is useful. For further details of how this notion works, see Wada (2008, 2010b).

Japanese does not have an exact equivalent of this "person" system, but has the dedicated term for private self (i.e. *zibun* 'self'), which serves to divide persons into self and others; this characterizes Japanese as private-self centered (see Hirose (2013) for fuller discussion).[32]

The second argument proposed by Hirose concerns the position of the speaker's deictic viewpoint in indirect speech.[33] As discussed in his studies (e.g. Hirose (1995, 2000)), indirect speech is a quotation of private expression; by using indirect speech, the reporter conveys the thought or mental representation (i.e. private expression) by the original speaker (or thinker) to the addressee. Take English (15), for example.

(15)   John said that Mary was sick in bed.

The original speaker's (i.e. John's) utterance may be "Mary is sick in bed" or "My girlfriend feels cold in bed." In either case, the reporter reduces the original speaker's utterance to the level of his/her private expression, i.e. his/her thought of Mary being sick in bed, and conveys the private expression to the addressee.

On this basis, Hirose provides the following generalization about the speaker's involvement in indirect speech: in both English and Japanese indirect-speech complements, their semantic content is associated with the perspective of the original speaker as private self because this environment is a quotation of private expression, whereas the deictic viewpoint for the form choice is attributed to the reporter as public self in English because of its public-self centeredness, but to the original speaker as private self in Japanese because of its private-self centeredness.[34] This generalization can explain

---

[32] The Japanese language certainly has a variety of terms referring to persons, such as *watasi* 'I (female or formal),' *boku* 'I (male, informal),' *ore* 'I (male, vulgar),' but they are not "personal" in the same sense as with West European languages. Yamaguchi (2009) calls the Japanese words referring to persons "person pronouns," distinguishing them from *personal* pronouns.

[33] This is also true of free indirect speech (represented speech).

[34] Hirose himself points out cases where, especially in English, the perspective of the reporter as public self is superimposed on that of the original speaker as private self with respect to the semantic content of indirect-speech complements (Hirose (1995, 1997)), as in (i):

(i)   John said that Mary is sick in bed.

Sentence (i) can be an indirect-speech version of (ii).

(ii)   John said, "Mary is sick in bed."

This might appear to violate the statement about the speaker's perspective with respect to the semantic content of indirect speech mentioned in the main text, but it does not. This is because in this example, the reporter still reports the private expression ascribed to the original speaker, and besides makes his/her own

the fact that to refer to the same past situation in the indirect-speech complement, English requires the past tense form *was*, as in (15) above, whereas Japanese requires the so-called present tense form *yokoninatteiru* 'lie,' or the *-ru* form in our terminology, as in (16).

(16) Tooru-wa Yooko-wa   byooki-de   yokoninat-tei-ru   to   it-ta.
      Toru-TOP Yoko-TOP   be sick-CF lie-STAT-NON-A   QUOT say-ANT
      'Toru said that Yoko was sick in bed.'

In English the past tense form is chosen because the complement situation is construed as occurring in the past relative to the deictic viewpoint of the reporter as public self. In Japanese, on the other hand, the *-ru* form is chosen because the complement situation is interpreted as occurring in the present relative to the deictic viewpoint of the original speaker as private self.

5.2. *Situation-Construal Tier, Situation-Report Tier, and Interpersonal-Relationship Tier*

We can now consider the three tiers in Hirose's model, i.e. the situation-construal tier, the situation-report tier, and the interpersonal-relationship tier. Hirose (2013:5) defines them as follows: in the situation-construal tier, "the speaker as private self construes the situation, forming a thought about it"; in the situation-report tier, "the speaker as public self reports or communicates his construed situation to the addressee"; and in the interpersonal-relationship tier, "the speaker as public self construes and considers his interpersonal relationship with the addressee." In the light of the integration patterns of the three tiers, he tries to explain typological characteristics of languages in a principled way. In particular, he argues that in English the situation-construal tier is in the default (i.e. normal) case integrated with the situation-report tier, both of which are separate from the interpersonal-relationship tier, but in Japanese the situation-construal tier is normally separate from the situation-report tier, which is integrated with the interpersonal-relationship tier. The integration of the situation-report tier and the situation-construal tier in English reflects its public-self centeredness because it guarantees the level of communication with the addressee; the separation of the situation-construal tier from the situation-report tier in Japanese reflects its private-self centeredness

---

judgment about the truth of the complement situation at his/her time of utterance, i.e. the time of the report. This is the mechanism of the superimposition in question. In this case, there are two perspectives activating with respect to the semantic content, and because of the public-self centeredness of English the perspective of the reporter as public self is given priority over that of the original speaker as private self and thus "foregrounded," though the latter perspective is still "active" in the "background." Therefore, the present tense form *is* is chosen to show that the complement situation still obtains in the present relative to the perspective of the reporter at the time of the report. See Wada (2001a) and Vandelanotte (2009) for more detailed explanation of the tense-interpretation mechanism.

because it does not guarantee such a level from the start, but only involves the level of the speaker's construal.

This characterization of the two languages automatically explains their many differences in language use. One such difference is as follows: an utterance, even without elements explicitly indicating addressee-orientedness (e.g. *you know, I tell you*), is normally interpreted as a communication act in English, as in (17), but such an utterance is normally interpreted as a thought expression in Japanese, as in (18), where without the addressee-oriented sentence-final particle *yo*, the sentence is normally interpreted not as addressed to others, but as the speaker's thought in his/her mind or the oral expression of it, i.e. his/her private expression.

(17) It's raining.
(18) Ame-da   (yo).
      rain-COP   SFP
      'It's raining.'

Since the situation-report tier is integrated with the situation-construal tier in English, uttering an English sentence automatically involves addressee-orientedness and is thus regarded as a communication (i.e. public expression) act. This is the default case in English. By contrast, since the situation-construal tier is separate from the situation-report tier in Japanese, uttering a Japanese sentence does not automatically involve addressee-orientedness and is thus regarded merely as a thought expression in the default case; to make it a communication act, addressee-oriented expressions such as *yo* 'I tell you' must be added if addressee-orientedness is not clear from the context or intonation.

Let us consider another difference, taken this time from Wada (2008), which can also be explained by the characterization in question. It concerns the fact that English has the subjunctive, i.e. the grammatical system for representing the speaker's mental attitude (though it is fairly obsolete now), whereas Japanese does not have such a grammatical system. Because English is a public-self centered language and the situation-report tier is integrated with the situation-construal tier, the situation construed by the speaker is in the default case reported from the perspective of the same speaker as public self. Therefore, to indicate that the relevant level of expression is a private expression (i.e. a thought expression or mental representation), a public-self centered language like English needs to have a special device to mark it explicitly; hence such a language has the subjunctive mood, i.e. a grammatical system for that purpose. In contrast, because Japanese is a private-self centered language and the situation-construal tier is separate from the situation-report tier, the relevant level of expression is usually construed as a private expression. Hence a grammatical device like the subjunctive mood is unnecessary in

Japanese. Instead, Japanese usually needs addressee-oriented (i.e. public) expressions to mark the relevant level of expression as a public expression, as we saw above.

6. The Association between the Unified Model of Tense and Modality and the Three-Tier Model of Language Use

We are now in a position to see how the unified model of tense and modality is motivated and supported by (and thus gives support to) the three-tier model of language use. To give the blueprint of my analysis, I will first associate the two models and present the "associated model." The English version and the Japanese version are respectively schematized in Figure 6 and Figure 7 below.
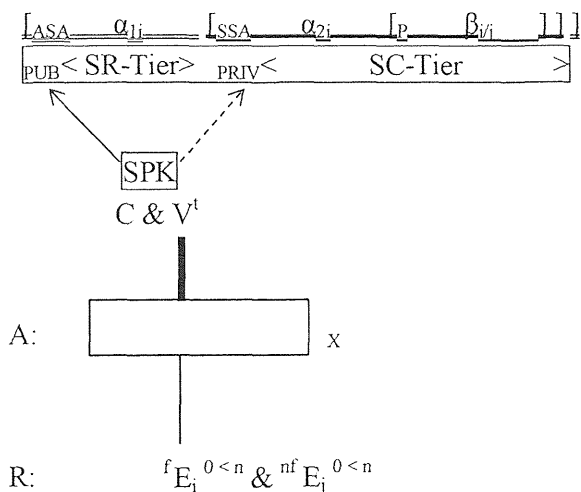
Fig. 6: A Unified Model of Tense and Modality Associated with the Three-Tier Model (English Version)
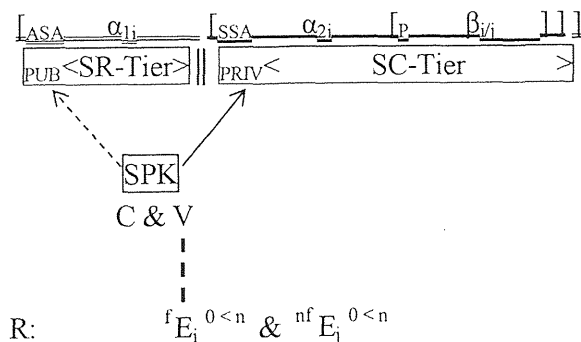
Fig. 7: A Unified Model of Tense and Modality Associated with the Three-Tier Model (Japanese Version)

Here, I focus on those parts of the model that we did not touch on in Figure 5 in section 4. One major characteristic of this associated model is that the SA (speaker's attitude) domain is divided into the ASA (addressee-oriented speaker's attitude) and the SSA (situation-oriented speaker's attitude) domains. Elements belonging to the ASA domain (symbolized by α_1), including Searle's illocutionary forces or points, are linked with the situation-report (SR) tier, whereas those belonging to the SSA domain (symbolized by α_2) and those belonging to the P (proposition) domain (symbolized by β) are linked with the situation-construal (SC) tier. The double and the bold underlines indicate elements in the situation-report tier and elements in the situation-construal tier, respectively. In both English and Japanese, situations described by finite verbs (predicates) can be associated with any of the three domains, i.e. the ASA, the SSA, and the P domains (the relation is represented by index i), whereas those described by non-finite verbs (predicates) are associated only with the P domain (the relation is represented by index j). The fact that the situation-report tier and the situation-construal tier are surrounded by one single box in Figure 6 and elsewhere in this article means that they are integrated with each other in English. The double vertical line in Figure 7 and elsewhere in this article indicates that the situation-construal tier is separate from the situation-report tier in Japanese. The symbols PUB and PRIV represent the public self (i.e. the subject of communication act or situation report) and the private self (i.e. the subject of thought expression or situation construal), respectively. The arrows extending from SPK to PUB or PRIV indicate that the speaker involved is interpreted either as public self or private self; the solid arrow represents the default interpretation pattern, whereas the dashed arrow shows a marked interpretation pattern, which requires certain conditions for it to be the case. As is inferable from the statements above, this model reflects the public-self centeredness of English and the private-self centeredness of Japanese.

Using this associated model, we can give a "deeper" or broader explanation for (a) the modal phenomena treated within the framework of my unified model of tense and modality (in section 4), (b) some differences of indirect speech acts between English and Japanese, and (c) some temporal phenomena that have been explained within my tense theory. For convenience's sake, I will consider the linguistic phenomena concerning modality and those concerning tense separately.

6.1. Field of Modality

We start by showing that the modal phenomena treated within the framework of the unified model of tense and modality are explained more systematically and comprehensively in the associated model. For this purpose, we need to note first that elements belonging to the SA domain in (8) above are further divided into those belonging to the situation-construal tier and those belonging to the situation-report tier, as shown in

Figures 6 and 7. The SA domain in the situation-report tier contains elements reflecting addressee-oriented speaker's attitudes (ASA elements) and is named the ASA domain; the speaker's aspect relevant to this domain is the public self. The SA domain in the situation-construal tier contains elements reflecting situation-oriented speaker's attitudes (SSA elements) and is dubbed as the SSA domain; the speaker's aspect relevant to this domain is the private self. Elements belonging to the P domain (P elements) belong to the situation-construal tier.[35] On these bases, I argue that as basic structures, English utterances include ASA elements in addition to SSA and P elements, whereas Japanese utterances do not include ASA elements, but consist only of SSA elements and P elements. Reorganizing hypothesis (8) in this way in terms of the three-tier model motivates our claim made in section 3 that the SA domain includes the level of Verstraete's (2001) "modal performativity," which corresponds now to the SSA domain. This result is crucial, especially for Japanese, because the combination of SSA elements and P elements is a "basic unit" in it. By contrast, because speech act theory is dedicated to communication acts (cf. Searle (1969:16)) and therefore the illocutionary force is assumed to belong to the ASA domain, the theory cannot guarantee the level of the speaker's attitude toward the situation (i.e. the SSA domain). In what follows, I will show how the associated model can explain the modal phenomena in English and Japanese.

6.1.1. English Cases

Let us start with English examples. The point here is that due to the integration of the situation-report tier with the situation-construal tier in English, uttering a sentence automatically indicates situation report in the default case. With this in mind, consider (19) first:

(19) John may come.

May in (19) is usually interpreted as expressing modality of possibility, a type of epistemic modality. This is the speaker's mental attitude when he or she construes the situation of John's coming. Therefore, this modality is an SSA element, belonging to the situation-construal tier. Since the situation-report tier is normally integrated with the situation-construal tier in English, uttering sentence (19) is normally accompanied by an ASA element, i.e. an element belonging to the situation-report tier. Unless otherwise

---

[35] The tropic, the neustic, and the phrastic components stated in Lyons (1977) appear to correspond respectively to what I call the ASA, the SSA, and the P domains. However, there are some differences. For example, in Lyons's system subjective epistemic modality "qualifies" the tropic component, but not the neustic component, whereas it is the opposite with objective epistemic modality; my associated model allows elements belonging to the ASA and the SSA domains to be all "subjective," so it makes irrelevant the problem of which domain is to be qualified in an utterance.

specified or implied, the speaker's attitude toward the situation is conveyed "untouched" to the addressee. The statements thus far are schematically represented in (20):

(20) $[_{ASA}$ possibility $[_{SSA}$ possibility $[_P$ John come $]]$ $]$

| < SR-Tier > < | SC-Tier | > |

In this article, the structure concerning what type of tier is piled up in making or interpreting a given utterance is called the "tier structure." The verb (predicate) in the P domain, when schematically represented, is represented in the base form because tense-dedicated elements like A-morphemes or R-morphemes are omitted (they are related to the field of tense). It should be stressed again that the modality of possibility accompanying the situation in question is represented in the SSA domain of the situation-construal tier, and the existence of the situation-report tier, by default, causes the same epistemic modality to occupy the ASA domain and accompany the message being conveyed to the addressee.[36]

I turn now to an unmodalized sentence like (21):

(21) Mary is a spy.

As we saw above, in my theory of modality, unmodalized sentences are accompanied by assertive modality in the default case.[37] In terms of the tier structure, such sentences are interpreted not only as implying the speaker's attitude toward the situation (i.e. assertion), but also as conveying the same attitude to the addressee. The point is schematized in (22):

(22) $[_{ASA}$ assertion $[_{SSA}$ assertion $[_P$ Mary be a spy $]]$ $]$

| < SR-Tier > < | SC-Tier | > |

In interpreting this sentence, assertion is inserted into the SSA domain and the same modality (i.e. assertion) as an ASA element is interpreted as being conveyed to the addressee because there is no indication otherwise in this case.

---

[36] Although Depraetere (2010, 2012) also takes the position that one single modal allows different meanings (interpretations) at more than one different level, her version differs from ours. For instance, she allows *can* to express possibility as context-independent semantic meaning, epistemic and root possibility as context-dependent semantic meanings, and illocutionary forces as pragmatic meanings. Therefore, this is related in some respects to, but different from, our distinction based on the three-tier model.

[37] In Searle's (1969, 1979) speech act theory, both the possibility as ASA element expressed by *may* in (19) and the assertion as ASA element conveyed in (21) can be reduced to "assertive" (an illocutionary point) in his terminology. I leave to future research a detailed description of the association between my ASA elements and Searle's illocutionary forces or points.

Now, I will consider a case where the ASA element of a given utterance is different from the SSA element, as illustrated in (23):

(23)   Mary is a spy, isn't it?

This utterance, consisting of the same situation described in (21) and a tag question, is assumed to have a falling tone on the tag question. Such a tag question is used to confirm that the speaker's remark is correct; it is taken as expressing the speaker's attitude "confirmation" directed to the addressee. The speaker's attitude toward the situation of Mary being a spy in this utterance is the same as that in (21); it is an assertion. Therefore, the utterance of (23) is schematically shown in terms of the tier structure, as in (24):

(24) $[_{ASA}$ confirmation $[_{SSA}$ assertion $[_{P}$ Mary be a spy $]]$ $]$
< SR-Tier > <	SC-Tier >

Here, the assertion (a type of epistemic modality) still underlies the confirmation (a type of speech act) at the situation-report tier. In other words, the assertion and the confirmation coexist in the ASA domain. When the ASA element which differs from the SSA element and the ASA element which is the same as the SSA element coexist in the ASA domain, it is natural that the former type is "foregrounded," because if it is not the case, there is no point in introducing that type of element in a communication act. Hence the confirmation is "featured" in the ASA domain; the assertion is "backgrounded."

Let us consider another case, where a so-called indirect speech act is brought about. I argue that the situation-report tier involved in an English sentential utterance enables us to induce indirect speech acts easily. Consider (25):

(25)   You must visit us when you come to the United States.

In this utterance, the speaker intends to show that the addressee will be welcome if the latter comes to the States, while the former construes the situation as a must in his/her mind (i.e. the obligation occupying the SSA domain).[38] By letting the addressee know it is an obligation (which now occupies the ASA domain), the speaker intends to get rid of the hesitation to visit him/her from the addressee. Given the contextual information indicating that the addressee is welcomed, together with the ASA element in question, the notion of invitation is derived in the situation-report tier. Because the invitation is an

---

[38] Cases like (25) and (27b) are examples of Searle's (1979) "indirect speech acts." Here, I am arguing only that the integration of the situation-report tier and the situation-construal tier in English "triggers" indirect speech acts; I am not discussing in detail how they are induced. I leave it to future research. For a possible analysis of this mechanism, see Searle (1979:Ch. 2).

ASA element different from the SSA element (i.e. obligation), this notion is featured. Therefore, the tier structure of (25) can be presented in such a way that the SSA element is considered to be an obligation and the ASA domain is interpreted as occupied by an invitation. This is schematized in (26):

(26) [$_{ASA}$ invitation   [$_{SSA}$ obligation [$_P$ you visit us when you come to the US]]]
< SR-Tier > <	SC-Tier	>

What is important here is that since the situation-report tier is integrated with the situation-construal tier in English, uttering sentences is ready for conveying ASA elements, and to this extent indirect speech acts are more likely to be induced in English, especially with sentences with modals. This is not the case with Japanese, as we will see later.

Let us next consider how *will*-sentences are analyzed in our associated model. Observe (27):

(27)　a.　　Akane will play the flute at tomorrow's concert.
　　　b.　　The Duty Officer will report for duty at 0700 hours.　　(Leech (2004:88))

In (27), *will* is usually interpreted as expressing predictive modality. Thus, in (27a), normally the speaker makes a prediction about the situation of Akane's playing the flute at tomorrow's concert (associated with the SSA domain) and is interpreted as conveying the same mental attitude (i.e. prediction) to the addressee (associated with the ASA domain). Therefore, sentence (27a) is, in terms of the tier structure, schematized below:

(28) [$_{ASA}$ prediction   [$_{SSA}$ prediction [$_P$ Akane play the flute]]]
< SR-Tier > <	SC-Tier	>

However, if sentence (27a) is interpreted, based on the contextual information, as expressing the subject's volition to play the flute at tomorrow's concert, then all the linguistic elements in (27a) belong to the P domain and thus assertive modality (the unmarked type of speaker's mental attitude) is inserted into the SSA domain. Unless otherwise implied by the context, the same modality is at the same time interpreted as an ASA element, which is triggered by the integration of the situation-construal and the situation-report tiers. The above statements are schematized in (29):

(29) [$_{ASA}$ assertion   [$_{SSA}$ assertion [$_P$ Akane will play the flute]]]
< SR-Tier > <	SC-Tier	>

Sentence (27b), on the other hand, is usually interpreted as conveying an instruction or order because it is usually uttered in a special context like the one in which the commander talks to his/her men. Therefore, while its SSA element is a prediction about the situation of the duty officer's report in the future, this special context, together with the prediction being conveyed to the addressee (which occupies the ASA domain), forces an instruction or order to be derived in the ASA domain of the situation-report tier as a featured ASA element. This is diagrammed in (30):

(30) [$_{ASA}$ instruction/order [$_{SSA}$ prediction [$_P$ The D.O. report for duty]]]

| < | SR-Tier | > < | SC-Tier | > |
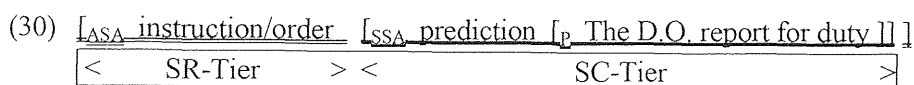
Here again, the integration of the two tiers induces an indirect speech act like this easily.

Finally, I will consider briefly how the treatment of the speaker's attitude elements in conditional sentences proposed by Wada (2011b) is reanalyzed in our associated model.[39]

(31) a. If it rains tomorrow, the match will be cancelled.

(Haegeman and Wekker (1984:45))

b. If it will rain tomorrow, we might as well cancel the match now.

(Haegeman and Wekker (1984:48))

The *if*-clause in (31a) and that in (31b) correspond to what I call "semantically deficient" and "semantically self-contained" clauses, respectively.[40] With respect to (31a), as we have seen, the composite proposition consisting of the P elements of the *if*-clause and the main clause falls under the "scope" of the predictive modality expressed by the *will* (i.e. an SA element) in the main clause. In our associated model, the sentence is analyzed in such a way that the prediction and the composite proposition both belong to the situation-construal tier and the same modality as an ASA element is inserted into the ASA domain of the situation-report tier by default. This insertion is triggered by the tier structure of English. The result of the analysis of (31a) is figured in (32).

---

[39] See Shizawa (2011) for comparisons of English and Japanese conditional sentences, especially what he calls speech-act conditionals, in terms of Hirose's dichotomy between public self and private self.

[40] Semantically deficient *if*-clauses and semantically self-contained *if*-clauses correspond respectively to the protasis of "predictive conditionals" and that of "non-predictive conditionals" in Dancygier (1998). Predictive conditionals are conditional sentences in which the tense of the conditional clause is "backshifted—a tense is backshifted when "the time marked in the verb phrase is earlier than the time actually referred to" (Dancygier (1998:37))—and the protasis and the apodosis have a (direct) cause-effect relationship with each other. Non-predictive conditionals are conditional sentences in which the tense form of the conditional clause is not "backshifted" and the protasis and the apodosis are relatively independent of each other (Dancygier (1998:61)).

(32) [$_{ASA}$ prediction [$_{SSA}$ prediction [$_P$ it rain tomorrow → the match be cancelled]]]

< SR-Tier > <	SC-Tier >

Note that the arrow in (32) indicates a causal relation.

A conditional sentence with a semantically self-contained *if*-clause like (31b), on the other hand, was analyzed in such a way that this type of *if*-clause consists of the SA and P domains and is thus equal to an independent or main clause in terms of the semantic decomposition of a sentential utterance in (8) above (see Wada (2011b)). It is often pointed out in the literature that this conditional clause is typically regarded as an echo conditional (in English). In this case, we can say, using Hirose's terminology, that the prediction expressed by *will* in the *if*-clause is attributed to the original speaker as private self and its semantic content is regarded as his/her private expression.[41]

Within our associated model, this phenomenon is reanalyzed in the following manner. Although both clauses have the SA and P domains, the *if*-clause has an SA element only in the situation-construal tier (i.e. an SSA element), whereas the main clause has two SA elements, i.e. an SSA and an ASA element. This indicates that although the *if*-clause and the main clause express independent statements (i.e. the propositional content with an SSA element), sentence (31b) as a whole constitutes one speech act (occupying the ASA domain), and it is the situation-report tier of the main clause that is responsible for the speech act. Since the conditional clause has "the inherently non-performative character" (Verstraete (2001:1519); cf. also Dancygier (1998)), it is not the case that the conditional clause piles up the situation-report tier in its tier structure. Therefore, the same mental attitude as the SSA element of the main clause, i.e. modality of weak possibility, is inserted into the ASA domain of the situation-report tier; that modality as an ASA element is conveyed to the addressee by the same speaker as public self. The result of the analysis of (31b) is schematized in (33):

(33) [$_{ASA}$ weak possibility [[$_{SSA}$ prediction [$_P$ it rain tomorrow]] => [$_{SSA}$ weak

< SR-Tier > <	SC-Tier >

possibility [$_P$ we cancel the match now]]]]

< SC-Tier >

The double arrow here (and elsewhere in this article) indicates that the left-side situation is a "trigger" to induce the right-side situation by inference. The semantic content of the *if*-

---

[41] In this connection, Dancygier (1998:120) states that the prediction expressed by *will* in the protasis of non-predictive conditionals represents the hearer's (addressee's) perspective and the speaker conveys a speech act "justified against the background of a prediction which only the hearer can make." In the case of echo conditionals, the hearer is taken as the original speaker. Thus, our statement in the main text is compatible with this characterization of Dancygier's.

clause (a private expression) serves as a trigger for the speaker as private self to conclude by inference that there is a weak possibility of the cancellation of the match, and the same speaker as public self conveys that weak possibility to the addressee.

6.1.2. Japanese Cases

I will now move on to observations of Japanese examples. As we saw above, Japanese is a private-self centered language and the situation-construal tier is normally separate from the situation-report tier.[42] This suggests that in the default case the speaker's mental attitude as an SSA element is not automatically conveyed to the addressee: an ASA element is not automatically inserted into the ASA domain because the situation-report tier does not "coexist" with the situation-construal tier from the start. Therefore, to make the construed situation a situation report, the speaker usually must add to it one or more addressee-oriented (i.e. public) expressions, i.e. elements which indicate explicitly that the situation-report tier is added to the tier structure in question.

With this in mind, let us analyze Japanese sentences. Consider (34):

(34) a.   Yooko-ga   yatteku-ru   daroo.
          Yoko-NOM   come-NON-A   will
          'Yoko will come.'

      b.   Yooko-ga   yatteku-ru   daroo   yo.
          Yoko-NOM   come-NON-A   will   SFP
          'Yoko will come.'

Sentence (34a) is, in the default case, interpreted as the speaker's monologue, i.e. merely as expressing the construed situation with his/her mental attitude toward it (prediction in this case). To convey it to the addressee, the speaker is expected to add an addressee-oriented expression like *yo* 'I tell you' and make it clear that his/her utterance is a situation-report, i.e. communication act, as in (34b). The separation of the situation-construal tier from the situation-report tier in Japanese normally invites us to regard (34a) as consisting only of elements belonging to the situation-construal tier, as figured in (35) below.

---

[42] In the three-tier model of language use, the situation-report tier in Japanese is integrated with the interpersonal-relationship tier, and as a result, in situation report, the speaker always has to pay due attention to who is talking to whom. Thus, even if the same content is reported, the speaker is supposed to use its polite form, as in *Watakusi-ga mairimasu* 'I will go (polite),' when talking to higher ranking or older people, while he can use a non-polite form, as in *Boku-ga iku* 'I will go (male, non-polite),' when talking to his friends or lower ranking people. The interpersonal-relationship tier is out of the scope of this article, so I will not touch on it unless necessary.

(35) [$_{SSA}$ prediction [$_P$ Yoko come]]

```
┌─────────────────────────────────┐
│ <          SC-Tier           >  │
└─────────────────────────────────┘
```

In the default case (especially cases where there is no evidence to show that the utterance is a communication act), the *daroo* 'will' in (34a) is interpreted as expressing the speaker's prediction about the situation that serves only as an SSA element.

By contrast, as illustrated in (34b), when an addressee-oriented expression (i.e. an element that guarantees the involvement of the situation-report tier) is added to the utterance normally interpreted as a situation construal, the whole utterance is interpreted as containing both the situation-construal tier and the situation-report tier, and thus, the speaker's mental attitude (i.e. prediction) is inserted into the ASA domain of the situation-report tier as an ASA element. Only after recognizing the addition of the situation-report tier can the addressee find the utterance a communication act for the first time and know that the predicted situation (i.e. the proposition "Yoko come" with the speaker's prediction) is conveyed to him/her. This is schematically represented in (36).

(36) [$_{ASA}$ prediction   [$_{SSA}$ prediction [$_P$ Yoko come]]]

```
┌──────────────┐ ┌───────────────────────────┐
│ < SR-Tier  > │ │ <         SC-Tier       >  │
└──────────────┘ └───────────────────────────┘
```

What is important is that in Japanese only when addressee-oriented expressions (i.e. public expressions) are added or the addressee-orientedness is clear from the context can the situation-report tier be connected to the situation-construal tier.

Next, we will show how unmodalized sentences are analyzed.

(37) a. Tooru-wa   Yooko-o   aisi-tei-ru.
        Toru-TOP   Yoko-ACC   love-STAT-NON-A
        'Toru loves Yoko.'

   b. Tooru-wa   Yooko-o   aisi-tei-ru   nda.
        Toru-TOP   Yoko-ACC   love-STAT-NON-A   SFP
        'Toru loves Yoko.'

In Japanese, like in English, independent or main clauses without any explicit modal expression can be interpreted as accompanied by assertive modality, i.e. the unmarked mental attitude of the speaker. The tier structure of (37a) is, in the default case, schematically represented, as in (38).

(38) [$_{SSA}$ assertion [$_P$ Toru love Yoko]]

```
┌─────────────────────────────────┐
│ <          SC-Tier           >  │
└─────────────────────────────────┘
```

If an addressee-oriented expression like *noda* 'it is that' (a marker of explanation) is added to sentence (37a), as shown in (37b), then the situation-report tier is connected to the situation-construal tier and therefore assertive modality as an ASA element is inserted into the ASA domain of the connected situation-report tier, as shown in (39).[43]

(39) $[_{ASA}$ assertion $[_{SSA}$ assertion $[_P$ Toru love Yoko $]]$]
$<$ SR-Tier $>$ $||$ $<$ SC-Tier $>$
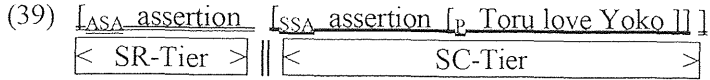
The addition of the situation-report tier to the tier structure in question enables the addressee to find that the asserted situation (i.e. the proposition "Toru love Yoko" with the speaker's assertion) is conveyed to him/her.

The difference in tier structure between English and Japanese in the default case leads us to be able to explain straightforwardly why in Japanese, unlike in English, a sentence containing *daroo* 'will' (expressing predictive modality) when its subject is second-person usually cannot be interpreted as conveying an instruction/order, as in (40a), or a sentence containing a modal of obligation like *nakerebanaranai* 'must' usually cannot be interpreted as conveying an invitation, as in (40b).

(40) a. (Anata-wa)   konban   hatizi-ni   hookokusu-ru   daroo.
        you-TOP   tonight eight o'clock-at report-NON-A   will
        '(Lit.) You will report at 8 o'clock tonight.'

     b. Nihon-o   otozure-ru   toki-wa   (anata-wa)   wareware-o
        Japan-ACC   visit-NON-A when-TOP   you-TOP   us-ACC
        tazune   nakerebanaranai.
        visit   must
        '(Lit.) When visiting Japan, you must visit us.'

Indirect speech acts like instructions/orders or invitations are ASA elements in our associated model. Since the tier structure of Japanese utterances normally does not involve the situation-report tier, addressee-oriented expressions are needed in order to add the situation-report tier to it. This implies that in Japanese sentential utterances, unlike in their English counterparts, the speaker's mental attitude as an SSA element is not inserted into the ASA domain of the situation-report tier automatically. Inferring an ASA element that is a different type of speaker's attitude from the SSA element imposes more burden on the addressee than receiving the same attitude as the SSA element. Therefore, the

---

[43] The marker of explanation *nda* in (37b) is a variant of *noda*. Here, we are considering *noda* as a public expression. In this connection, Ikarashi (2012) suggests a possibility of treating both *noda*-sentences as public expressions and those as private expressions from a unified point of view.

former process is more difficult in Japanese based only on the contextual information and without any addressee-oriented (i.e. public) expression that guarantees the addition of the situation-report tier (cf. (25) and (27b)). For this reason, sentences (40a) and (40b) cannot receive an interpretation in which an ASA element indicating an instruction/order or invitation occurs in the ASA domain of the situation-report tier; the addition of expressions indicating such speech acts is necessary for such an interpretation.[44]

Finally, let us consider semantically self-contained conditional clauses (cf. Wada (2011b)), i.e. cases where elements belonging to the SA domain occur in Japanese conditional clauses (cf. also Shizawa (2011)). Consider the following examples:[45]

(41) a.   Yooko-ga   Kitahara-to   kekkonsuru   kamosirenai   nara,
          Yoko-NOM   Kitahara-with   marry   may   if
          Tooru-wa   doona-ru   no   ka.
          Toru-TOP   become-NON-A   COMP   Q
          'If Yoko may marry Kitahara, what will become of Toru?'

     b.   Yooko-ga   Kitahara-to   kekkonsuru   kamosirenai   nara,
          Yoko-NOM   Kitahara-with   marry   may   if
          Tooru-wa   doona-ru   no   ka nee.
          Toru-TOP   become-NON-A   COMP   Q SFP
          'If Yoko may marry Kitahara, what will become of Toru?'

As is clear from what we have discussed, sentence (41a) as it stands does not include the situation-report tier in the default case and is interpreted as a private expression; without further contextual information or an addressee-oriented (i.e. public) expression, this question is regarded merely as an expression of the speaker's thought, not addressed to anybody. On the other hand, as shown in sentence (41b), the addition of the public expression *nee* 'I tell you' to sentence (41a) brings about the addition of the situation-report tier to its tier structure; as a result, the sentence is, in terms of the tier structure, analyzed in the same way as its English counterpart (e.g. (31b)). The schematization of the tier structure of (41b) is as follows:

---

[44] In this respect, too, our associated model is superior to speech act theory. Because it should be the case in speech act theory that both English and Japanese utterances equally involve illocutionary acts, i.e. acts that consist of illocutionary forces (or points) and propositions, the theory alone cannot explain the differences concerning indirect speech acts considered in the main text.

[45] As pointed out in Masuoka (1991, 1997), because clauses containing *nara* (i.e. *nara*-clauses) allow modality (i.e. elements belonging to the SA domain in my terminology) to occur in them, they are used as examples of Japanese conditional clauses for the present purpose.

(42) [$_{ASA}$ question [[$_{SSA}$ possibility [$_P$ Toru love Yoko]] => [$_{SSA}$ question [$_P$

| < SR-Tier > ‖ | < | SC-Tier | > |

what become of Toru]]]]

| < | SC-Tier | > |

Here, the modality of possibility in the *nara*-clause (i.e. the modality expressed by *kamosirenai* 'may') is attributed to the present speaker as private self (or the original speaker as private self in the case of echo conditionals), and the semantic content of the *nara*-clause causes the speaker (as private self) to infer that of the main clause. The speaker's mental attitude which is the same type as the SSA element, i.e. question, is inserted into the ASA domain of the situation-report tier as an ASA element because of the addition of the public expression *nee*, and as a result, the whole sentence is interpreted as a communication act. Note, in passing, that if we remove the part corresponding to the situation-report tier from (42), we will obtain the tier structure for (41a).

6.2. Field of Tense

Finally, I will show that the analysis of temporal phenomena in my tense theory can be treated in a more motivated or "deeper" way by our associated model. My characterization of the different tense systems of English and Japanese (in my previous studies) is, in terms of our associated model, restated as follows: since English is a public-self centered language and a sentential utterance of it involves the situation-report tier (to which the public self is relevant) in its tier structure by default, it has tense forms with the A(bsolute tense)-component in its tense system; by contrast, since Japanese is a private-self centered language and a sentential utterance of it does not involve the situation-report tier in its tier structure by default, it does not have tense forms with the A-component, but only has those with the R(elative tense)-component, in its tense system.[46]

First of all, let me show how our associated model legitimates our statement that English has tense forms with A-morphemes, but Japanese does not. As is extensively discussed in Hirose (2013), the grammatical notion of "person" is a linguistic sign showing that the utterance with such a notion is interpreted as a situation report, or public expression, which involves the public self and presupposes addressee-orientedness (see also the relevant statements in section 5.1 above). This observation motivates my claim that English has A-morphemes.[47] The reason is as follows: since A-morphemes are tense morphemes integrated with the notion of person, using absolute tense forms (tense forms with the A-component occupied by A-morphemes), or English finite forms, naturally

---

[46] What is mainly added to the restatement is the statement about the tier structure.

[47] This way of motivation in connection with the difference in tier structure between English and Japanese has not been provided in my previous studies and thus is a merit of our associated model.

reflects the perspective of the public self (the role of the speaker as communicating subject). In this way, the involvement of public self in the basic tier structure of English sentential utterances gives a reason for the existence of the A-morpheme in English.

On the other hand, Japanese does not have the grammatical notion "person" comparable to that of English. There is no grammatical motivation for the existence of tense morphemes reflecting the perspective of the public self. Therefore, in Japanese, the existence of A-morphemes is not motivated, and even finite forms do not have the A-component in the tense structure; as a result, the tense forms have R-morphemes, i.e. tense morphemes that do not change according to person. In this way, the exclusion of public self from the basic tier structure of Japanese sentential utterances gives a reason for the non-existence of the A-morpheme in Japanese. Since there is no fusion of the speaker's t-viewpoint and consciousness at speech time in the Japanese tense system, the identification of the base point in time for the choice of tense forms, finite or non-finite, depends on the type of the linguistic environment. For this reason, especially in the linguistic environment in which the subject of situation report or communication (i.e. public self) is not directly relevant, such a base point in time (i.e. the time at which the speaker's viewpoint is situated depending on the nature or characteristics of the relevant linguistic environment) can be a local event time rather than speech time.

Now, let us reanalyze the temporal phenomena in indirect speech in terms of our associated model. Consider, again, the different patterns in the (finite) tense-form choice in English and Japanese indirect-speech complements (observed by Hirose) discussed in section 5.1, which are illustrated again in (43) below.

(43) a.   John said that Mary was sick in bed.   (= (15))
     b.   Tooru-wa   Yooko-wa   byooki-de   yokoninat-tei-ru   to   (= (16))
          Toru-TOP   Yoko-TOP   be sick-CF   lie-STAT-NON-A   QUOT
          it-ta.
          say-ANT
          'Toru said that Yoko was sick in bed.'

The reason stated there was that English is a public-self centered language and thus requires the deictic viewpoint of the reporter as public self (i.e. the subject of the whole sentence in (43a)) for its tense-form choice, whereas Japanese is a private-self centered language and thus requires the deictic viewpoint of the original speaker as private self (i.e. Toru in (43b)) for its tense-form choice. Hirose's three-tier model can further motivate this characterization.[48] The integration of the situation-construal tier with the situation-report tier in English "features" the public self even in this linguistic environment, i.e. a

---

[48] This is also the case with free indirect speech, which offers another piece of evidence for this view.

quotation of private expression (to which the perspectives of the reporter as public self and the original speaker as private self are relevant) and hence shows the public-self centeredness. The separation of the situation-construal tier from the situation-report tier in Japanese excludes the "interference" of the public self and "features" the private self in the same environment and hence shows the private-self centeredness.

However, this view itself does not explain the difference between the deictic patterns of the tense-form choice and those of the deictic-adverb choice in indirect speech. Consider (44), for example:

(44) a. John said that he would arrive yesterday. (Comrie (1985:116))

b. Tooru-wa   Yooko-wa   sakuban   tuk-u   to   it-ta
   Toru-TOP   Yoko-TOP   last night arrive-NON-A QUOT   say-ANT
   noni....
   but
   'Toru said that Yoko would arrive last night, but ....'

Not only in English but also in Japanese, the base point for choosing the deictic adverbs (i.e. *yesterday* and *sakuban* 'last night') is located at speech time (i.e. the deictic center of the real time line). On the other hand, the base point for choosing the tense forms is different in English and Japanese: it is speech time in English, but the time of the original utterance in Japanese.

Our associated model (involving my tense theory) can explain this fact systematically. Let us first confirm that the semantic content of the indirect-speech complement itself consists only of the situation-construal tier (i.e. a quotation of private expression); as Hirose shows, in (45) and (46) the modal expressions in the complement clauses are attributed to the original speaker as private self (cf. Brecht (1974)).[49]

(45) a. John says that the news is probably true. (Hirose (1995:234))

b. John thinks that Mary may possibly be a spy. (Hirose (1995:34))

---

[49] Here, we are assuming cases where the reporter conveys the original speaker's private expression neutrally. As Hirose himself points out, in English as a public-self centered language, unlike in Japanese as a private-self centered language, it is sometimes possible that the perspective of the reporter as public self is superimposed on the semantic content of the indirect-speech complement. In our terms, this is analyzed as follows. The sentences in (45) in the text, for instance, can be interpreted in such a way that the modality attributed to the reporter as public self which is the same type as the modality attributed to the original speaker as private self is superimposed on the original speaker's private expression, and is taken as an ASA element, i.e. an element belonging to the situation-report tier. See also note 34. I leave a detailed investigation of this case to future research.

(46) Taroo-wa   ame   nitigainai   to   omot-tei-ru.[50]   (Hirose (1997:10))
Taro-TOP   rain   must   QUOT   think-STAT-NON-A
'Taro thinks that it must be raining.'

However, this linguistic environment is also related to both of the perspectives of the original speaker as private self and the reporter as public self. This means that ASA elements (or public expressions) related to the perspective of the reporter as public self can be superimposed on the semantic content. It is safe to say that the deictic adverbs in question intrinsically include reference to speech time, i.e. the time at which is situated the reporter's deictic viewpoint (i.e. a part of the reporter's perspective). Taking these observations together, we can account for why both English and Japanese deictic adverbs can be attributed to the deictic viewpoint of the reporter, as in (44) above.

Now, let us explain the different patterns of the finite tense forms in English and Japanese. Since English finite forms contain A-morphemes, the base point in time for their choice is speech time, i.e. the time of the report (to which the reporter as public self, or the subject of situation report, is relevant); since Japanese finite forms do not contain A-morphemes, their tense-form choice is subject to the nature or characteristics of the relevant linguistic environment, so that the private-self centered nature of the indirect-speech complement (i.e. the original speaker's private expression) forces the viewpoint for choosing Japanese finite forms to be on the time of the utterance of the original speaker as private self, or the subject of situation construal. In this way, our associated model can provide a "deeper" or more motivated account of the differences of deictic phenomena in indirect speech than my compositional tense theory alone or the characterization of the two languages by the dichotomy between public-self centeredness and private-self centeredness alone.

Our associated model can also explain why even in English (a public-self centered language), if the linguistic environment is concerned only with the private self (i.e. the subject of situation construal), then the tense form to be chosen is a relative one. Consider (47):

(47) a.   I saw [Akiko play the violin].
     b.   I want [to go to Italy].

Semantically, the bracketed parts describe no more than the situation the speaker saw or desires. Syntactically, they constitute an embedded clause. In our model, they are taken

---

[50] I use the verb of thinking in the Japanese case because the complement clause of the verb of saying in Japanese can be ambiguous between a quotation of direct speech and one of indirect speech (see Hirose (1997) for details).

as the linguistic environments in which only the P domain of the situation-construal tier attributed to the speaker as private self is relevant (these environments do not contain the SA domain because modal expressions cannot occur in them). This characterization prevents the speaker as public self from being involved because these environments are "private worlds" attributed to the private self. Therefore, absolute tense forms (i.e. tense forms with A-morphemes representing the public self's viewpoint) cannot be used here.

7. Conclusion

In this article, I have first presented the unified model of tense and modality by combining my compositional tense theory and the developed version of the modality theory utilized to supplement it, and then shown how it is related with and motivated by (and lends support to) the three-tier model of language use proposed by Hirose (2013). Although the associated model we have developed is just a grand design and needs more verification for details, I hope to have shown that it has the potential to give a more comprehensive analysis of temporal and modal phenomena in English and Japanese from a unified point of view. Our next task is to extend the model not only to cover more data of the temporal and modal phenomena of the two languages but also to be able to explain those phenomena of other languages such as German, Dutch, and French, e.g. the phenomena that have already been analyzed in my previous studies (Wada (2002, 2010b, 2013)). I will reserve them for another occasion.

REFERENCES

Allen, Robert (1966) *The Verb System of the Present-Day American English*, Mouton, The Hague.

Benveniste, Emile (1971) *Problems in General Linguistics*, trans. by Mary Elizabeth Meek, University of Miami Press, Coral Gables, Florida.

Brecht, Richard (1974) "Deixis in Embedded Structures," *Foundations of Language* 11, 489-518.

Caudal, Patrick (2012) "Pragmatics," *The Oxford Handbook of Tense and Aspect*, ed. by Robert Binnick, 269-305, Oxford University Press, Oxford.

Close, Reginard A. (1977) "Some Observations on the Meaning and Function of Verb-Phrases Having Future Reference," *Studies in English Usage: The Resources of a Present-Day English Corpus for Linguistic Analysis*, ed. by Wolf-Dietrich Bald and Robert Ilson, 125-156, Peter Lang, Frankfurt.

Comrie, Bernard (1976) *Aspect*, Cambridge University Press, Cambridge.

Comrie, Bernard (1985) *Tense*, Cambridge University Press, Cambridge.

Dancygier, Barbara (1998) *Conditionals and Prediction: Time, Knowledge, and Causation in Conditional Constructions*, Cambridge University Press, Cambridge.

Declerck, Renaat (1991) *Tense in English: Its Structure and Use in Discourse*, Routledge, London.

Declerck, Renaat (1997) When-*Clauses and Temporal Structure*, Routledge, London.

Declerck, Renaat (in cooperation with Susan Reed and Bert Cappelle) (2006) *The Grammar of the English Verb Phrase Volume I: The Grammar of the English Tense System: A Comprehensive Analysis*,

Mouton de Gruyter, Berlin/New York.

Depraetere, Ilse (2010) "Some Observations on the Meaning of Modals," *Distinctions in English Grammar, Offered to Renaat Declerck*, ed. by Bert Cappelle and Naoaki Wada, 72-91, Kaitakusha, Tokyo.

Depraetere, Ilse (2012) "Tense in Sentences with Modal Verbs," *The Oxford Handbook of Tense and Aspect*, ed. by Robert Binnick, 989-1019, Oxford University Press, Oxford.

Duffley, Patrick (1992) *The English Infinitive*, Longman, London.

de Haan, Ferdinand (2012) "Evidentiality and Mirativity," *The Oxford Handbook of Tense and Aspect*, ed. by Robert Binnick, 1020-1046, Oxford University Press, Oxford.

Haegeman, Lilliane and Herman Chr. Wekker (1984) "The Syntax and Interpretation of Futurate Conditionals in English," *Journal of Linguistics* 20, 45-55.

Harder, Peter (1996) *Functional Semantics: A Theory of Meaning, Structure, and Tense in English*, Mouton de Gruyter, Berlin/New York.

Hasegawa, Yoko and Yukio Hirose (2005) "What the Japanese Language Tells Us about the Alleged Japanese Relational Self," *Australian Journal of Linguistics* 25, 219-251.

Hirose, Yukio (1995) "Direct and Indirect Speech as Quotations of Public and Private Expression," *Lingua* 95, 223-238.

Hirose, Yukio (1997) "Hito o Arawasu Kotoba to Syoo-oo (Words of Reference to Persons and Anaphora)," *Sizi to Syoo-oo to Hitei (Reference, Anaphora, and Negation)*, ed. by Minoru Nakau, 1-89, Kenkyusha, Tokyo.

Hirose, Yukio (2000) "Public and Private Self as Two Aspects of the Speaker: A Contrastive Study of Japanese and English," *Journal of Pragmatics* 32, 1623-1656.

Hirose, Yukio (2002) "Viewpoint and the Nature of the Japanese Reflexive *Zibun*," *Cognitive Linguistics* 13, 357-401.

Hirose, Yukio (2013) "Deconstruction of the Speaker and the Three-Tier Model of Language Use," *Tsukuba English Studies* 32, 1-28.

Hirose, Yukio and Yoko Hasegawa (2010) *Nihongo kara Mita Nihonzin: Syutaisei no Gengogaku (Japanese People as Seen from the Japanese Language: The Linguistics of Subjectivity)*, Kaitakusha, Tokyo.

Huddleston, Rodney (1977) "The Futurate Construction," *Linguistic Inquiry* 8, 730-736.

Ikarashi, Keita (2012) "Abudakusyon ni Motozuita '*No(da)*'-Bun ni Kansuru Nitiei Hikaku (A Contrastive Study of *No(da)* Constructions and Related Expressions in English from the Perspective of Abduction)," *JELS* 29, 52-58.

Janssen, Theo A.J.M. (1994) "Tense in Dutch: Eight 'Tenses' or Two Tenses?," *Tense Systems in European Languages*, ed. by Rolf Thieroff and Joachim Ballweg, 93-118, Niemeyer, Tübingen.

Janssen, Theo A.J.M. (1996) "Tense in Reported Speech and Its Frame of Reference," *Reported Speech*, ed. by Theo A.J.M. Janssen and Wim van der Wurff, 237-259, John Benjamins, Amsterdam/Philadelphia.

Kamp, Hans and Uwe Reyle (1993) *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic, and Discourse Representation Theory*, Kluwer Academic,

Dordrecht.

Klein, Wolfgang (1992) "The Present Perfect Puzzle," *Language* 68, 525-552.

Klein, Wolfgang (1994) *Time in Language*, Routledge, London.

Langacker, Ronald (2008) *Cognitive Grammar: A Basic Introduction*, Oxford University Press, Oxford.

Leech, Geoffrey (2004) *Meaning and the English Verb*, 3rd Edition, Longman, London.

Lyons, John (1977) *Semantics*, Two Volumes, Cambridge University Press, Cambridge.

Masuoka, Takashi (1991) *Modarithi no Bunpoo (The Grammar of Modality)*, Kurosio, Tokyo.

Masuoka, Takashi (1997) *Fukubun (Complex Sentence)*, Kurosio, Tokyo.

Nakau, Minoru (1994) *Ninti Imiron no Genri (Principles of Cognitive Semantics)*, Taishukan, Tokyo.

Nuyts, Jan (2005) "The Modal Confusion: On Terminology and the Concepts behind It," *Modality*, ed. by Alex Klinge and Henrik Høeg Müller, 5-38, Equinox, Oakville.

Palmer, Frank R. (2001) *Mood and Modality*, 2nd Edition, Cambridge University Press, Cambridge.

Rathert, Monika (2012) "Adverbials," *The Oxford Handbook of Tense and Aspect*, ed. by Robert Binnick, 237-268, Oxford University Press, Oxford.

Searle, John (1969) *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge.

Searle, John (1979) *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge University Press, Cambridge.

Shizawa, Takashi (2011) *Form, Meaning, and Discourse: The Semantics and Pragmatics of Conditional Constructions in English and Japanese*, Doctoral dissertation, University of Tsukuba.

Vandelanotte, Lieven (2009) *Speech and Thought Representation in English: A Cognitive-Functional Approach*, Mouton de Gruyter, Berlin/New York.

Verstraete, Jean-Christophe (2001) "Subjective and Objective Modality: Interpersonal and Ideational Functions in the English Modal Auxiliary System," *Journal of Pragmatics* 33, 1505-1528.

Wada, Naoaki (2001a) *Interpreting English Tenses: A Compositional Approach*, Kaitakusha, Tokyo.

Wada, Naoaki (2001b) "Eigo no Kanryookei, Nihongo no Kanryookei Sootoohyoogen no Zikankoozoo to Teiziten Fukusirui tono Kyookisei (The (In)Compatibility of the Perfect Form with Adverbials of Definite Time Position in English and Japanese), *Gengo Kenkyuu* 119, 77-110.

Wada, Naoaki (2002) "A Comparative Study of the English Present Perfect and the German Perfekt: With a Special Reference to Their Differences in Co-Occurrence with Adverbials Referring to a Definite Time," *English Linguistics* 19, 335-365.

Wada, Naoaki (2008) "Kootekiziko Tyuusinsei no Doai to Seioosyogo no Hoo-/Zisei-Gensyoo no Sooi (The Degree of Public-Self Centeredness and Differences of Mood/Tense Phenomena in Western Languages)," *Kotoba no Dainamizumu (The Dynamism of Language)*, ed. by Yu-ichi Mori, Yoshiki Nishimura, Susumu Yamada and Mitsuaki Yoneyama, 277-294, Kurosio, Tokyo.

Wada, Naoaki (2009a) "'Uti' no Siten/'Soto' no Siten to Zisei Gensyoo: Nitieigo Taisyookenkyuu ('Inside' Viewpoint/'Outside' Viewpoint and Tense Phenomena: A Contrastive Study of Japanese and English)," *'Uti' to 'Soto' no Gengogaku (The Linguistics of 'Uti' and 'Soto': Inside and Outside)*, ed. by Atsuro Tsubomoto, Naoko Hayase and Naoaki Wada, 249-295, Kaitakusha, Tokyo.

Wada, Naoaki (2009b) "The Present Progressive with Future Time Reference vs. *Be Going To*: Is Doc Brown Going Back to the Future Because He Is Going to Reconstruct It?," *English Linguistics* 26, 96-131.

Wada, Naoaki (2010a) "On the Distinction of English Past Tenses," *Distinctions in English Grammar, Offered to Renaat Declerck*, ed. by Bert Cappelle and Naoaki Wada, 37-61, Kaitakusha, Tokyo.

Wada, Naoaki (2010b) "Genzai Kanryookei no Arawasu Imi Han-i no Hensen to Sii Ken-in: Eirango Hikaku Kenkyuu (The Semantic-Range Shift of Present Perfect Forms and C-Gravitation: A Comparative Study of English and Dutch)," *Bungei Gengo Kenkyuu: Gengo-hen* (*Studies in Language and Literature: Language*) 58, 75-112.

Wada, Naoaki (2011a) "On the Mechanism of Temporal Interpretation of *Will*-Sentences," *Tsukuba English Studies* 29, 37-61.

Wada, Naoaki (2011b) "Nitieigo Zisei Gensyoo no Taisyoo Gengogakuteki Bunseki: Jookensetu o Zireikenkyuu to site (A Contrastive-Linguistic Analysis of Tense Phenomena in Japanese and English: A Case Study of Conditional Clauses)," *Bungei Gengo Kenkyuu: Gengo-hen* (*Studies in Language and Literature: Language*) 60, 69-120.

Wada, Naoaki (2013) "Eigo to Furansugo no Miraihyoogen no Hikaku (Some Comparisons between English and French Future Expressions)," *Bungei Gengo Kenkyuu: Gengo-hen* (*Studies in Language and Literature: Language*) 63, 107-146.

Wada, Naoaki (to appear) "On the So-Called Future-Progressive Construction," *English Language and Linguistics* 17, 391-414.

Watanabe, Junya (2004) *Furansugo no Syookosei no Imiron* (*The Semantics of Evidentiality in French*), Soobi-Syuppansya, Tokyo.

Yamaguchi, Haruhiko (2009) *Meiseki na In-yoo/Sinayaka na In-yoo* (*Explicit Quotation and Flexible Quotation*), Kurosio, Tokyo.

Faculty of Humanities and Social Sciences
University of Tsukuba
e-mail: wada.naoaki.gb@u.tsukuba.ac.jp