TBTA Tutorial Lesson 2: Concepts, Ontology, and Semantic Representations

(video: TBTA-2A The Ontology)

2.0 Introduction

This lesson will introduce you to TBTA's concepts, ontology, and semantic representations. The concepts, ontology, and semantic representations are all supplied for you, but you need to become familiar with the semantic complexity levels of the concepts, the organization of the ontology, and the format of the semantic representations. Later you'll learn how to add your own concepts to the ontology, and how to build your own semantic representations.

2.1 Concepts

A concept is a structured idea. For example, English speakers are familiar with the concept of "a written message that is sent or delivered to someone." This particular concept has been lexicalized in English with the word *letter*. Some concepts are semantically simple, while other concepts are semantically complex. For example, concepts such as *good, bad, big, small, think, know, want, feel,* etc., are all semantically simple. Other concepts such as *saunter* (to walk in a slow and relaxed way so that you look confident or proud), *waddle* (to walk taking short steps, swaying from side to side like a duck), *wade* (to walk in shallow water), etc., are semantically complex. Cross-linguistic research indicates that semantically simple concepts are much more likely than semantically complex concepts to have lexical equivalents in other languages.

Natural semantic metalanguage (NSM) theorists such as Anna Wierzbicka have proposed that there is a small set of innate concepts which are present in every language. They claim that these innate concepts are semantically very simple, and can be used to define every word in every language. They call these innate concepts "semantic primitives." A complete list of the semantic primitives is shown in Table 2-1. If semantic representations were developed using only these semantic primitives, then the problem of lexical mismatch between languages would be eliminated. However, developing semantic representations using only the semantic primitives is unwieldy, and the intended message would be distorted. Therefore semanticists use "semantic molecules" which are semantically more complex than the primitives, but still semantically simple. One of the foundational principles of NSM states that semantically simple concepts may be identified by searching for words that are frequently used in the definitions of other words. The developers of the Longman Dictionary of Contemporary English used a computer program to scan through the definitions of their words and identify the words that are used most frequently in their definitions. The result was a list of approximately 3,000 words which they call the

Longman Defining Vocabulary. TBTA has elected to use the Longman Defining Vocabulary as its semantic molecules, and the vast majority of the concepts in TBTA's semantic representations are from this vocabulary list.

Substantives	I, YOU, SOMEONE, PEOPLE/PERSON, SOMETHING/THING
Mental Predicates	THINK, KNOW, WANT, FEEL, SEE, HEAR
Speech	SAY, WORD
Actions, Events, and Movement	DO, HAPPEN, MOVE
Existence	THERE IS
Life	LIVE, DIE
Determiners	THIS, THE SAME, OTHER
Quantifiers	ONE, TWO, SOME, ALL, MANY/MUCH
Evaluators	GOOD, BAD
Descriptors	BIG, SMALL
Time	WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME
Space	WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE
Interclausal Linkers	BECAUSE, IF
Clause Operators	NOT, MAYBE
Metapredicate	CAN
Intensifier, Augmentor	VERY, MORE
Taxonomy, Partonomy	KIND OF, PART OF
Similarity	LIKE

Table 2-1 List of NSM Semantic Primitives (Goddard 1998:58)

For convenience, TBTA's semantic representations use concepts that have been lexicalized by English, and English labels are used to represent these concepts. Therefore the concept of "a written message that is sent or delivered to someone" will be represented by LETTER[1]. But the English word *letter* has other senses or meanings. For example, the word *letter* can also refer to a character in an alphabet. In order to distinguish the written message from the character in an alphabet, the written message will be represented in the semantic representations by LETTER-A, and the alphabetic character will be represented by LETTER-B. So *letter* is an English word and it has at least two senses or meanings. LETTER-A has just one meaning; LETTER-A represents the concept of "a written message that is sent to someone." Because TBTA's concepts are very narrowly defined, you'll be able to easily map these concepts to target words or constructions.

When discussing objects such as *letters*, the various senses are generally quite clear. When discussing events, attributes or relations, the various senses are more difficult to distinguish. For example, consider the English word *open*.

[1] Throughout these tutorials, capital letters will be used to distinguish concepts from English words.

This word has many different senses, two of which are: 1) to remove the lid or cover of a container so that things may be put in or taken out of the container, and 2) to move a door, window or gate of some type in a certain way so that people or things may move in or out of a room, building, city, etc. Most languages will probably require two different verbs for these two senses of *open*, so the ontology has OPEN-A and OPEN-B. Additionally in English we can say things like *John opened his eyes*, *John opened his mouth*, or *John opened a book*. These senses of *open* are quite distinct from OPEN-A and OPEN-B so your target language will probably require different verbs for these senses of *open* as well. However, it's impractical to define an additional sense for each object that is being opened. Therefore concepts have been chosen which represent the major senses of each event, attribute and relation in the ontology. Throughout the semantic representations OPEN-A has been used when a person opens a container of any kind, OPEN-B is used when opening a door, window, or gate, OPEN-C is used when opening a book, OPEN-D is used when a person or animal opens its eyes, and OPEN-E is used when a person or animal opens its mouth. If your language requires different target verbs for opening a letter versus opening a box, then you'll need to write one or more collocation correction rules to handle these particular cases. Those rules will be described in a subsequent tutorial.

2.2 The Ontology

Ontology is the philosophical study of the nature of existence. When used in the context of natural language generators, the word "ontology" refers to a structured collection of concepts. Most NLGs have a restricted list of concepts which are permitted in the system's semantic representations, and these concepts are generally categorized in some way within the ontology.

To see the ontology in TBTA, select Ontology and Lexicon in the main menu, and then select Ontology, Objects to Nouns. You'll see a screen with two tables. The yellow table contains all of the concepts in the ontology, and the blue table will eventually contain your target lexicon. A small section of the ontology is shown below in Figure 2-1.

	Concept Stems	Senses	Mappings	English Glosses
554	leprosy	A		the skin disease
555	letter	A		a letter, note
556	letter	B		a letter of the alphabet
557	Levi	A		Jacob's son

Figure 2-1. The Ontology

The ontology is organized into seven semantic categories: 1) objects, which are realized by English nouns, 2) events, which are realized with English verbs, 3) object attributes, which are realized by adjectives, 4) event attributes, which are realized by adverbs, 5) relations, which are realized by English prepositions, 6) conjunctions and 7) particles. In order to keep these tutorials simple, these semantic categories will be called by their standard syntactic equivalents: nouns,

verbs, adjectives, adverbs, adpositions, conjunctions and particles. Similarly object phrases will be called noun phrases, propositions will be called clauses, etc.

As you work with the concepts in the ontology, be sure to carefully read the gloss provided for each concept. The concepts are always represented with English words, but the actual concept may be somewhat different than your idea of what that word generally means. For example, open the dropdown in the upper left corner of the ontology screen and select Relations. The first seventeen relations don't have lexicalized English equivalents so they're given descriptive names. An example of a few of these relations follows:

7. ITERATION-A
[Clause [NP JOHN] [VP HIT] [NP PETER] [NP ITERATION [AdjP 3] TIMES]]
John hit Peter three times.

10. NAME-A
[NP MAN [NP NAME MARK]] …
A man named Mark …

14. QUANTITY-A
[Clause [NP JOHN] [VP BUY] [NP SILVER [NP QUANTITY [AdjP 3] KILOGRAM]]]
John bought three kilograms of silver.

17. TITLE-A
[NP ISAIAH [NP TITLE PROPHET]] …
Isaiah, the prophet, …

As indicated by the name, relations signal some type of semantic relationship between two objects or events. Scroll downward until you come to the relation WHEN-B. Its gloss is "when-because + event specifies the time and reason for the main event." An example of this relation occurs in the sentence *When the sink overflowed, I called a plumber*. The *when* in that sentence signals both a temporal relationship between the two events as well as a reason relationship. This is a very different sense of *when* than in the sentence *John was six when he learned to read*. The *when* in that sentence signals just a temporal relationship. Other languages may treat these two senses of *when* differently. The important point here is that you should look at the gloss of each concept to make sure that you understand it before you deal with it in your target language. The ontology includes twenty-one senses of BE. No language will have twenty-one different types of stative clauses, but you should make sure that you understand each individual concept before you enter a target word or rule that will be triggered by that concept.

The concepts in the ontology have also been divided into five levels of semantic complexity[2]:

[2] You can see these five levels by clicking the Setup button in the upper left corner of the ontology screen, and then looking at the upper right corner of the Setup dialog.

- Semantic Primitives
- Semantic Molecules
- Complex Concepts with Insertion Rules
- Complex Concepts without Insertion Rules
- Inexplicable Concepts

The first two levels are called “Semantic Primitives” and “Semantic Molecules,” and they were both described above. The semantic primitives have a blue background in the “Senses” cell, and the semantic molecules have a light yellow background in the “Senses” cell.

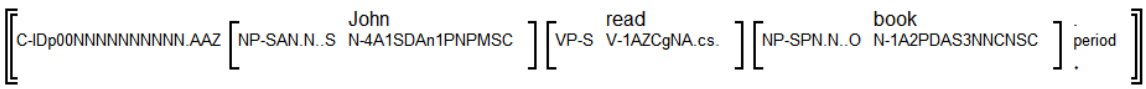
The next category is called Complex Concepts with Insertion Rules. These concepts never appear directly in the semantic representations, but they may be activated if your target language has a good lexical equivalent. For example, the concept BLIND-A is semantically complex so it never appears in the semantic representations; instead the relative clause *is not able to see* is used in the semantic representations. But if your language has a lexical equivalent for BLIND-A, you can activate the rule for this concept. Then all of the relative clauses *is not able to see* in the semantic representations will be automatically converted to BLIND-A. The process of activating the complex concept insertion rules will be described in a subsequent tutorial.

The complex concepts in the next category don’t have insertion rules, but the developers of TBTA are working to build complex concept insertion rules for them. The number of concepts in this level is constantly becoming smaller, and eventually there won’t be any concepts in this category.

The final level of semantic complexity is called “Inexplicable,” meaning either that these concepts can’t be adequately explicated using a simple phrase or relative clause, or that they’re completely inexplicable. All proper names and numbers are in this category, and many artifacts (e.g., *bow* and *arrow*) are also considered inexplicable.

2.3 The Semantic Representations

A semantic representation is any formal method of representing meaning. Linguists have proposed many different formats for semantic representations, but because TBTA deals with minority languages, a new format had to be developed that is specifically oriented toward these languages. These semantic representations have been developed using a controlled English based metalanguage[3] augmented by a feature system designed to accommodate a very wide variety of languages. For example, the semantic representation for the proposition *I should finish reading these books* is shown below in Figure 2-2.



[3] A metalanguage is any language that is used to describe or analyze another language. In a bilingual dictionary where Korean words are described in English, English is considered a metalanguage because it is being used to describe Korean.

Figure 2-2. Semantic Representation of *I should finish reading these books*.

These semantic representations are composed of concepts, features, and structures. As seen in Figure 2-2, there are three concepts in this proposition: JOHN-A, READ-A, and BOOK-A. The features in the semantic representation are indicated by the sequences of characters below each concept and next to each proposition and phrase boundary. The structure is indicated by the proposition boundaries and phrase boundaries. The concepts in the semantic representations have already been described, so the next two sections will discuss the features and structures.

2.3.1 The Feature System used in the Semantic Representations

In TBTA each semantic/syntactic category has a set of features that are specifically relevant to it. Each of these features has two parts: 1) a name and 2) a set of possible values. For example, one of the features associated with nouns is named Number, and the possible values are Singular, Dual, Trial, Quadrial, Plural, and Paucal. As you develop your lexicon and grammar, you'll define additional features that are relevant to your particular language. The features that you define will be assigned to specific syntactic categories, and they will consist of a name and a set of possible values.

To see a semantic representation in TBTA, click on Generator in the main menu, and Lexicon and Grammar Development. The large window contains the semantic representation of the current verse. You can see the reference for the current verse in the upper left corner of this screen. This is the screen where you will do the vast majority of your work, so you need to become familiar with it.

Features are displayed in the semantic representations as seen above in Figure 2-2. When you rest your cursor on the features in the Lexicon and Grammar Development screen, you'll see a popup explaining what each of those characters means. Shown below in Figure 2-3 is the same semantic representation that was shown above in Figure 2-2, but this time the popups explaining the features have also been included.

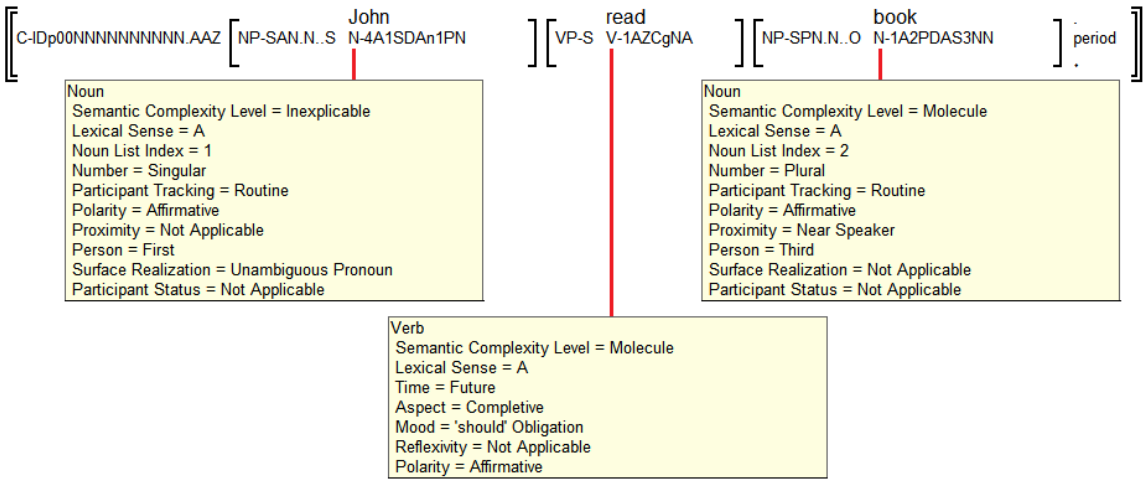


Figure 2-3. Semantic Representation with Popups Explaining the Features

For each concept, the first two features will always be Semantic Complexity Level and Lexical Sense, both of which were described in the previous sections. The other features are specific to the concept's semantic category. Phrases and clauses also have features, but there wasn't enough space in the figure above to include their popups.

As was mentioned above, Figure 2-3 shows the semantic representation for the sentence *I should finish reading these books*. The popup for the verb shows that the Aspect is Completive and the Mood is 'should'. So those two features are used to generate *should finish* in English. The Person feature below JOHN is First indicating that John is the speaker. There are no pronouns in the semantic representations, but every noun has a Person feature. The Number feature on BOOK is Plural and the Proximity feature is Near Speaker. So those two features are used to generate *these books*. You can now see how this semantic representation produces the English sentence *I should finish reading these books*.

Following is a brief discussion of the semantic/syntactic categories, their features and their values. To see a complete list of all the features and their values, click on Generator in the main menu, and then Lexicon and Grammar Development. In the lower right corner is a button labeled Feature Set. When you click that button, you'll see a dialog with two drop downs at the top. The left drop down contains all the syntactic categories, and the right drop down contains all the features for the selected syntactic category. The table on the dialog lists all the values for the selected feature.

1) Objects – Nouns

A. Noun List Index: 1, 2, 3, …

The Noun List Index feature is distinct from the other features. Your grammar will never directly access this feature. In other words, you won't write rules that say something like, "When the Noun List Index is 3, add the affix XYZ." The Noun List Index feature is used to determine whether two nominals are the same or different. For example, this feature is used to identify which noun in a relative clause is coreferential with the head noun. This feature is also used to determine if the subject of an object complement is coreferential with the subject of the matrix clause. For example, $JOHN_1$ WANT [$JOHN_1$ BUY CAR_2] represents *John wants to buy a car*. Similarly, $JOHN_1$ WANT [$MARY_2$ BUY CAR_3] represents *John wants Mary to buy a car*. In the first example where the subject noun in the matrix clause has the same index as the subject noun in the complement clause, you'll want a particular set of rules to apply; when the two nouns have a different index as in the second example, you'll want a different set of rules to apply. The Noun List Index feature is also useful when working with a clause chaining language that uses a switch reference system. To properly generate the Same Subject and Different Subject morphemes, you'll use the

Noun List Index feature. This feature will be discussed more thoroughly in the tutorial that describes pronoun rules.

B. Number: Singular, Dual, Trial, Quadrial, Plural, Paucal

Very few languages morphologically distinguish all six of these possible values. In a subsequent lesson you'll write rules that will collapse feature values so that your grammar only has to deal with the values that are relevant to your language. So if your language only distinguishes Singular and Plural, your grammar will only need to deal with Singular and Plural.

C. Participant Tracking: First Mention, Integration, Routine, Exiting, Restaging, Offstage, Generic, Interrogative, Frame Inferable

A few languages introduce participants onto the discourse stage in a two step process. The first time a participant is mentioned in a discourse, it is tagged as First Mention. The second time it's mentioned, it's tagged as Integration. For the remainder of the discourse, the nominal will be tagged as Routine. The last time a participant appears in a discourse, it will be tagged as Exiting. If a participant is brought back onto the discourse stage, it will be tagged as Restaging. An example of Generic is found in Adjectives 1:8, *John read many books*. Here the nominal BOOK is tagged as Generic. When a nominal is tagged as Interrogative, we generally want the English word *which* before it. An example is found in Clauses 1:97, *Which book did John read?* When a nominal is tagged as Frame Inferable, English inserts the definite article *the* as in *I bought a car last week. Unfortunately <u>the engine</u> had a problem*. In the second sentence ENGINE will be tagged as Frame Inferable rather than First Mention.

D. Polarity: Affirmative, Negative

Virtually all of the nouns in the semantic representations are tagged as Affirmative. When a noun is tagged as Negative, English places the word *no* before it as in *No man has climbed that mountain*. In this proposition MAN is tagged with Negative Polarity. Your language may form the semantic equivalent using an affirmative generic plural noun with a negative verb as in *Men have not climbed that mountain*.

E. Proximity: Not Applicable, Near Speaker and Listener, Near Speaker, Near Listener, Remote within sight, Remote out of sight, Temporally Near, Temporally Remote, Contextually Near, Contextually Near with Focus

This feature is used to generate the demonstratives *this* and *that*. Physical objects in direct quotes will be assigned the values Near Speaker and Listener, Near Speaker, Near Listener, Remote within sight, and Remote out of sight. Only temporal objects will be assigned the values Temporally Near and Temporally Remote. For example, we can talk about *this Christmas* or *that*

weekend. The objects that are tagged as Contextually Near or Contextually Near with Focus are those objects that aren't physically present, but are still on the discourse stage. For example, we can talk about *that man* even when his physical location is unknown.

F. Person: First, Second, Third, First & Second, First & Third, Second & Third, First & Second & Third

Many languages have a First Person Inclusive pronoun and a First Person Exclusive pronoun. First Person Inclusive is the equivalent of First & Second. First Person Exclusive is equivalent to First, and possibly also First & Third. The values Second & Third and First & Second & Third are included only for completeness; they will probably never occur in the semantic representations.

G. Surface Realization: Not Applicable, Always a Noun, Unambiguous Pronoun

The vast majority of the nouns in the semantic representations are tagged with a Surface Realization value of Not Applicable. When this is the case, your rules will decide which nouns should be realized by target nouns and which should be realized by pronouns. The value Always a Noun is only used with nouns that have a Person value of Third. When a noun is tagged as Always a Noun, you should probably generate a target noun. For example, John 1:1 says *In the beginning was the Word, and the Word was with God, and the Word was God*. In that verse you'll probably want all three occurrences of *Word* to be realized with a noun. Therefore those three instances of WORD are tagged as Always a Noun. No nouns in the semantic representations are marked with Unambiguous Pronoun. You'll need to write language specific rules that determine when pronouns should be used. When a rule decides that a noun should be realized with a pronoun, the rule should set this feature to Unambiguous Pronoun. Then a subsequent spellout rule can examine the nouns that are set to Unambiguous Pronoun and generate the proper surface forms.

H. Participant Status: Not Applicable, Protagonist, Antagonist, Major Participant, Minor Participant, Major Prop, Minor Prop, Significant Location, Insignificant Location

A few languages morphologically indicate protagonists and antagonists so they've been marked throughout the semantic representations. It's fairly common for languages to introduce significant participants, props and locations in a different way than the less significant participants, props and locations. For example, when a significant prop or location is first mentioned, it may be placed at the beginning of the clause. If your language doesn't make use of this information, you can hide this feature. Hiding irrelevant features will be discussed in a subsequent tutorial.

2) Events - Verbs

A. Time: Discourse, Present, Immediate Past, Earlier Today, Yesterday, 2 Days Ago, 3 Days Ago, A Week Ago, A Month Ago, A Year Ago, During Speaker's Lifetime, Historic Past, Eternity Past, Unknown Past, Immediate Future, Later Today, Tomorrow, 2 Days from Now, 3 Days from Now, A Week from Now, A Month from Now, A Year from Now, Unknown Future, Timeless

Time is the semantic equivalent of the syntactic term Tense. Many languages have multiple degrees of past tense and multiple degrees of future tense. Since those languages define their degrees differently, the divisions here are finer than the divisions made by any one particular language. Therefore every language will require rules to collapse these Time values into a smaller set. Most of the verbs in the semantic representations are tagged as Discourse; the other values are only used in direct speech. In many languages the type of discourse determines the general tense that is used. English uses past tense for narrative discourse, but that isn't universal. Your rules that generate tense will need to look at the Discourse Type feature that is on every clause.

Timeless tense always coincides with Gnomic aspect. Timeless tense and Gnomic aspect are used in sentences that are always true, e.g., *Oil floats on water*, *The sun rises in the east*, *God loves us*, etc. English encodes timeless tense and gnomic aspect with present tense, but again that isn't a universal.

B. Aspect: Unmarked, Habitual, Imperfective, Progressive, Completive, Inceptive, Cessative, Continuative, Gnomic

The aspects included here are the aspects that significantly affect the meaning of the text. Some of the more subtle aspects such as Perfective have not been included in this list because they don't significantly affect the generated text. Another common aspect, Perfect, has not been included in this list. Indo-European languages generally use Perfect aspect to signal flashback. Flashback is a Salience Band and Salience Bands are indicated on each proposition. Another notable aspect that is missing from this list is Iteration. In the semantic representations Iteration is indicated with either an adverbial phrase containing the concept REPEATEDLY, or an oblique NP containing the specific number of times the event was iterated. The vast majority of the verbs in the semantic representations are tagged with Unmarked aspect.

C. Mood: Indicative, Definite Potential, Probable Potential, 'might' Potential, Unlikely Potential, Impossible Potential, 'must' Obligation, 'should' Obligation, 'should not' Obligation, Forbidden Obligation, 'may' (permissive)

Mood indicates an event's relationship to reality. Indicative mood signals that the event did or will happen; all of the other moods indicate that the event hasn't actually happened. Examples of how English encodes each of these moods follow:

Definite Potential: *John will definitely walk to the store*.
Probable Potential: *John will probably walk to the store*.
'might' Potential: *John might walk to the store*.
Unlikely Potential: *John will probably not walk to the store*.
Impossible Potential: *John definitely will not walk to the store*.
'must' Obligation: *John must walk to the store*.
'should' Obligation: *John should walk to the store*.
'should not' Obligation: *John should not walk to the store*.
Forbidden Obligation: *John must not walk to the store*.
Permissive 'may': *John may walk to the store*. (For example, John's parents are giving him permission to walk to the store.)

D. Reflexivity: Not Applicable, Reciprocal, Reflexive

Reciprocal and Reflexive situations are marked on the verb even though the morphological encoding for these events generally appears in the object phrase. An example of a reciprocal clause is *We saw each other*. An example of a reflexive clause is *I saw myself*.

E. Polarity: Affirmative, Negative, Emphatic Affirmative, Emphatic Negative

Emphatic Affirmatives are encoded in English with something like *I will definitely read that book*. An example of an Emphatic Negative is *I definitely did not read that book*.

3) Object Attributes - Adjectives

A. Degree: Not Applicable, Comparative, Superlative, Intensified, Extremely Intensified, 'too,' 'overly', 'less', 'least'

Most adjectives in the semantic representations are tagged with a Degree value of Not Applicable. For example, in the sentence *John is tall*, and in the phrase *the tall man*, both occurrences of TALL would be tagged with a Degree value of Not Applicable. English examples for the other values of Degree follow:
Comparative: *John is taller than Steve*.
Superlative: *John is the tallest man in the room*.
Intensified: *John is very tall*.
Extremely Intensified: *John is extremely tall*.
'too,' 'overly': *John is too tall*.
'less': *This book is less important than that book*.
'least': *That book is the least important book*.

Not all languages have comparative constructions as in *John is taller than Steve*. The semantic equivalent in some languages may be something like *Steve is tall but John is very tall*. If your target language uses the latter construction,

you'll be able to write a rule that will appropriately restructure all sentences that contain comparative adjectives.

4) Event Attributes - Adverbs

A. Degree: Not Applicable, Comparative, Superlative, Intensified, Extremely Intensified, 'too', 'less', 'least'

Similar to Adjective Degree, most adverbs in the semantic representations are tagged with a Degree value of Not Applicable. For example, in the sentence *John walked quickly*, QUICKLY is tagged with a Degree value of Not Applicable. English examples of the other values for Degree are listed below.
Comparative: *John walked more quickly than Steve*.
Superlative: *John walked the most quickly*.
Intensified: *John walked very quickly*.
Extremely Intensified: *John walked extremely quickly*.
'too': *John walked too quickly*.
''less': *John walked less quickly than Steve*.
'least': *John walked the least quickly*.

Similar to the comparative adjectives, not all languages have comparative adverbs. Rather than saying *John walked more quickly than Steve*, some languages may say *Steve walked quickly, but John walked very quickly*. If that's the case in your language, you'll be able to write a rule to deal with those constructions. The semantic representations will never compare two events that are different. For example, the semantic representations will never say *John walked more quickly than Steve ran*. In that sentence one person's *walking* is being compared with another person's *running*. Although this is permissible in real language, it's not permissible in the semantic representations.

5) Relations - Adpositions

There aren't any features associated with relations.

6) Conjunctions

There aren't any features associated with conjunctions.

7) Particles

There aren't any features associated with particles.

8) Object Phrases - Noun Phrases

A. Type: Simple, First Coordinate, Last Coordinate, Coordinate

In TBTA's semantic representations there are four types of phrases: noun phrases, verb phrases, adjective phrases and adverb phrases. There are no prepositional phrases. In TBTA's semantic representations, prepositions occur in noun phrases as well as in adverbial clauses. All four phrase types have a feature called Type. In general the Type feature indicates how many phrases there are of a particular type. For noun phrases, the Type feature indicates how many noun phrases with a particular semantic role occur in that clause. For example, if a Patient NP is tagged as Simple, then it is the only Patient NP in that clause. In the sentence *John kicked the ball*, the NP containing BALL is tagged as Simple because it is the only Patient NP in the clause. If there are multiple patient NPs in a clause, then the first will be tagged with First Coordinate and the last will be tagged with Last Coordinate. If there are three or more NPs in a clause that all have the same semantic role, then the NPs that aren't first or last will each be tagged with Coordinate. So in the sentence *John kicked the ball, the bat, the base and the glove*, the NP containing BALL will be tagged as First Coordinate, the NPs containing BAT and BASE will both be tagged as Coordinate, and the NP containing GLOVE will be tagged as Last Coordinate.

B. Semantic Role: Agent, Patient, State, Source, Destination, Instrument, Addressee, Beneficiary, Not Applicable

Linguists have proposed many different sets of semantic roles. The set of semantic roles used in TBTA has intentionally been kept very small. Rather than using a large set of very descriptive and precise semantic roles, TBTA uses a small set of generic semantic roles. This makes it much easier for users to write rules that will convert the generic semantic roles into roles that are appropriate for their languages.

The term Agent is used as a generic semantic role to cover the roles traditionally labeled as agents, forces, and experiencers. The Patient semantic role generally is used for the most affected argument. The State semantic role is used almost exclusively in stative clauses. For example, in the sentence *John is a teacher*, the NP containing TEACHER is labeled as a State and the NP containing JOHN is labeled as Agent. The other semantic roles are self explanatory.

Even with a small set of semantic roles, the role assigned to a particular nominal may be debatable. You may not always agree with the semantic role assignments in the semantic representations, and in some cases the role assignments may seem somewhat arbitrary. For example, in the sentence *Mary named the baby John* (Noun Phrases 1:25), the NP containing MARY is the Agent, the NP containing BABY is the Patient, and the NP containing JOHN is a State. The precise identification of each nominal's semantic role isn't important. When generating text in another language, what's important is that the semantic representations label the various roles for each event consistently. One of the tools used during the development of the semantic representations guarantees that the semantic roles will be labeled consistently for each event. Therefore

you'll be able to write rules that will change the semantic roles to whatever is appropriate for your language.

In the main menu select Ontology and Lexicon, Ontology, Events to Verbs. When you look at the events in the ontology, there's a special column labeled Theta Grids. In that column there's a sequence of coded letters. To see what those letters mean, rest your cursor in that column for a particular event. You'll see a popup telling you which semantic roles are obligatory for that event, which are optional, and which never occur. For example, scroll down to ANSWER-A and rest the cursor in its Theta Grid cell. The popup will tell you that ANSWER-A always has an Agent NP, it has an optional Patient NP, and it always has a Patient Proposition. You'll then see a series of examples illustrating each of the possible situations. As you can see in the popup, there's an example for each possible case frame situation for ANSWER-A. You'll see this same information in a popup when you're looking at the semantic representation of a verse and you rest the cursor on a verb. For example, in the main menu select Generator, Lexicon and Grammar Development. Rest the cursor on any verb and you'll see a popup showing its case frame and examples of all its possible argument structures. To see this popup, make sure that you rest the cursor on the verb itself and not on its features.

9) Event Phrases - Verb Phrases

A. Type: Simple, First Coordinate, Last Coordinate, Coordinate

As was discussed under Noun Phrases, this feature indicates how many verb phrases are present in a particular clause. Since each proposition has only one verb, there are no examples of a coordinate VP in the semantic representations. But this feature is included because you may occasionally write rules that introduce additional VPs, and you'll need to know which VP occurs first, and which comes last.

10) Object Attribute Phrases - Adjective Phrases

A. Type: Simple, First Coordinate, Last Coordinate, Coordinate

Coordinate adjective phrases are very rare. You can see an example in Adjectives 1:12 which produces *That book is big and red*. The adjective phrase containing BIG is tagged as First Coordinate and the adjective phrase containing RED is tagged as Last Coordinate.

B. Usage: Attributive, Predicative

English adjectives may be used either attributively, as in the phrase *the tall man*, or they may be used predicatively as in the clause *John is tall*. Adjectives used attributively will always be in an adjective phrase that's embedded in a noun phrase. Adjectives used predicatively will always be in an

adjective phrase that's in a clause. An example of an attributive adjective can be found at Adjectives 1:1, *John read a good book*. An example of a predicative adjective can be found at Adjectives 1:11, *That book is too good*. When an adjective is used predicatively, the verb will always be BE-D. You may want to write transfer rules that will convert predicative adjectives to target verbs. If that's the case, your transfer rule should be triggered by BE-D. This operation will be described thoroughly in a subsequent tutorial.

11) Event Attribute Phrases - Adverb Phrases

A. Type: Simple, First Coordinate, Last Coordinate, Coordinate

Similar to the adjective phrases, coordinate adverb phrases are very rare. You can see an example in Adverbs 1:6 which produces *John walked quickly and carefully*. The adverb phrase containing QUICKLY is labeled First Coordinate and the adverb phrase containing CAREFULLY is labeled Last Coordinate.

12) Propositions - Clauses

A. Type: Independent, Coordinate Independent, Restrictive Thing Modifier, Descriptive Thing Modifier, Event Modifier, Participant, Patient, Attributive Patient

There are two general categories of clauses: independent and dependent. Independent clauses are able to stand alone and are labeled Independent in the semantic representations. There are no propositions in the semantic representations labeled Coordinate Independent. This value is included here because you'll eventually write rules that will combine simple clauses into more complex clauses when appropriate conditions are satisfied. The other clause types listed above are all dependent clauses. Restrictive Thing Modifier is the semantic name for a restrictive relative clause, e.g., *The man <u>that I saw</u> bought that book*. Descriptive Thing Modifier is the semantic name for a descriptive relative clause, e.g., *John, <u>who will be arriving later today</u>, will open the meeting*. In this sentence the listeners already know who John is; the relative clause adds extra information rather than identifying who John is. Event Modifier is the semantic equivalent of an adverbial clause. All adverbial clauses begin with a relation/preposition that signals the relationship between the main clause and the adverbial clause. A Participant proposition is a subject complement clause (e.g., *It pleased Mary that <u>John read that book</u>*.) English generally inserts the expletive *it* and then postposes the subject complement. A Patient proposition is an object complement (e.g., *John wanted <u>Mary to read a book</u>*.) An Attributive Patient proposition is an object complement for an adjective (e.g., *John is afraid <u>to read that book</u>*.)

B. Illocutionary Force: Declarative, Imperative, Content Interrogative, Yes-No Interrogative, 'let' Imperative

All of these values are self explanatory except the last one. ‘let’ imperatives only occur in the first chapter of Genesis. For example, in Gen. 1:3, *God said, “Let there be light*.”

C. Topic NP: Agent, Patient

Every clause is marked to signal the semantic role of the topicalized nominal. Generally the topic NP is the Agent; when this is the case, the clause is active. If the clause’s topic NP is the Patient, the clause is passive. Your target language may not have active and passive constructions, but every language has some method of emphasizing one particular nominal over the other nominals. The Topic NP feature indicates which nominal is to be emphasized. Real languages generally permit almost any NP to be topicalized, but TBTA’s semantic representations only permit the agent and patient NPs to be topicalized.

D. Speaker: Not Applicable, Adult Daughter, Adult Son, Angel, Animal, Boy, Brother, Crowd, Daughter, Demon, Disciple, Employee, Employer, Father, Girl, God, Government Leader, Government Official, Group of Friends, Holy Spirit, Husband, Jesus, King, Man, Military Leader, Mother, Prophet, Queen, Religious Leader, Satan, Servant, Sister, Slave, Slave Owner, Soldier, Son, Wife, Woman, Written Material to General Audience (letter,law,etc.)

Many languages use a system of honorifics when people speak to one another. For example, when parents talk to their children, they may use one set of pronouns or a particular verbal affix, but when children talk to their parents, the children may use a different set of pronouns or a different verbal affix. Similarly men may have one style of speaking, and women may have another. One culture may honor its priests, another culture may honor its warriors. Since it’s impossible to predict how or when a particular language will indicate honor or deference during direct speech, every direct quote in the semantic representations has been marked to indicate the general category of the speaker and the listener. For example, in Ruth 2:2 Ruth is talking to Naomi, her mother-in-law. Those sentences are direct quotes and they’re tagged with ‘Adult Daughter’ as the Speaker and ‘Mother’ as the Listener.

E. Listener

This feature has the same values as Speaker.

F. Speaker’s Attitude: Not Applicable, Neutral, Familiar, Endearing, Honorable, Derogatory, Friendly, Antagonistic, Complimentary, Anger, Rebuke

When two people who don't know each other are talking and there are no particular emotions involved, the speech will be tagged as Neutral. If the two people know one another, the speech will be tagged as Familiar if there are no particular emotions involved. The other speaker attitudes listed above are used when two people who know one another are speaking and the speaker has a particular attitude or emotion.

G. Speaker's Age: Not Applicable, Child (0 – 17 years old), Young Adult (18 – 24), Adult (25 – 49), Elder (50+)

In speech situations the ages of the speaker and listener are often relevant to the speech style that will be selected. This feature indicates the approximate age of the speaker, and the next feature indicates whether the listener is older or younger.

H. Speaker-Listener Age: Not Applicable, Older - Different Generation, Older - Same Generation, Essentially the Same Age, Younger - Different Generation, Younger - Same Generation

Using the example in Ruth 2:2 again, Ruth is talking to her mother-in-law. Ruth is one generation younger than her mother-in-law, so the Speaker-Listener Age feature is set to Younger – Different Generation.

I. Direct Speech Style: Not Applicable

This feature is intentionally left blank so that you can define the various speech styles that are relevant to your language. For example, Korean has six speech styles: Deferential, Polite, Blunt, Familiar, Intimate, and Plain. You can define whatever styles are used in your language. This process will be explained thoroughly in a subsequent tutorial.

J. Discourse Genre: Climactic Narrative Story, Episodic Narrative Story, Narrative Prophecy, Expository, Behavioral Hortatory, Behavioral Eulogy, Procedural, Persuasive, Expressive, Descriptive, Epistolary, Dramatic Narrative, Dialog

Every clause is marked to indicate its discourse genre. The discourse genres listed above come from *The Grammar of Discourse* by Robert E. Longacre, p. 10.

K. Notional Structure Schema: Not Applicable, Narrative-Exposition, Narrative-Inciting Incident, Narrative-Developing Conflict, Narrative-Climax, Narrative-Denouement, Narrative-Final Suspense, Narrative-Conclusion, Hortatory-Authority Establishment, Hortatory-Problem or Situation, Hortatory-Issuing of Commands, Hortatory-Motivation, Procedural-Problem or Need, Procedural-Preparatory Procedures, Procedural-Main Procedures, Procedural-

Concluding Procedures, Persuasive-Problem or Question, Persuasive-Proposed Solution or Answer, Persuasive-Supporting Argumentation, Persuasive-Appeal, Expository-Problem or Situation, Expository-Solution or Answer, Expository-Supporting Argumentation, Expository-Evaluation of Solutions

The most significant value of Notional Structure Schema is Narrative-Climax. Many languages have a special set of rules that apply only at the peak of a narrative discourse. Therefore each sentence is tagged to indicate its notional structure schema[4].

L. Salience Band: Not Applicable, Pivotal Storyline, Primary Storyline, Secondary Storyline, Script Predictable Actions, Backgrounded Actions, Flashback, Setting, Irrealis, Evaluation, Cohesive Material

In every discourse regardless of its genre, there is a mainline as well as various types of supportive material. The supporting material encodes progressive degrees of departure from the mainline[5]. Each language has its own techniques for distinguishing the mainline from the different types of supportive material. Additionally each language has its own methods of encoding each of the types of supportive material. For example, English uses perfect aspect to encode Flashback as in the sentence *John had already studied for the test*. Later you'll be able to write rules that apply to one or more particular salience bands.

M. Sequence: Not in a Sequence, First Coordinate, Last Coordinate, Coordinate

Similar to the NPs, there may occasionally be multiple subordinate propositions of the same type embedded within a proposition. For example, two relative clauses might modify the same noun, or one verb might have two object complements. These situations are rare, but when they occur, the values First Coordinate and Last Coordinate are used to determine the clause order.

N. Location in Paragraph: Not Applicable, First, Last, Discourse Title, Aperture, Finis, Footnote, Questionable Text

This feature is used to indicate section titles, footnotes, biblical passages which may not have been in the original manuscripts (e.g., Mark 16:9-20), etc.

O. Implicit Information: Not Applicable, Implicit Cultural Information, Implicit Situational Information, Implicit Historical Information, Implicit Background Information, Implicit Subactions

---

[4] For additional discussion of notional structure schema, see Longacre, Robert E., The Grammar of Discourse, Second Edition, chapters 1 and 2.

[5] For additional discussion of salience bands, see Longacre, Robert E., The Grammar of Discourse, Second Edition, p. 21.

Often implicit information must be made explicit in order for the texts to communicate their intended meanings. Five different types of implicit information have been identified, and whenever implicit information is added to the semantic representations, the information is tagged appropriately.

P. Alternative Analysis: Not Applicable, Primary Analysis, First Alternative Analysis, Second Alternative Analysis, Third Alternative Analysis, Fourth Alternative Analysis, Fifth Alternative Analysis

There are passages which even the best biblical scholars are uncertain of the intended meaning (e.g., Luke 23:31 *For if they do these things in a green tree, what will be done in the dry?)*. In these cases the semantic representations will include the best interpretations that scholars have proposed, and they will be marked to indicate that they are alternative analyses.

2.3.2 The Structures used in the Semantic Representations

The structural system in TBTA is generally very simple and clear. As you saw in Figure 2-1, all phrases are enclosed in single brackets, and all clauses are enclosed in double brackets. Noun phrases and adjective phrases may be embedded in noun phrases. Relative clauses are also embedded in noun phrases as shown below in Figure 2-4.

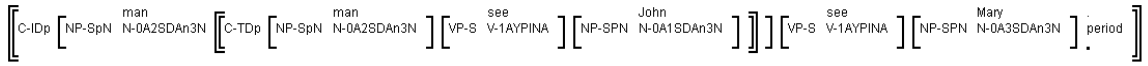


Figure 2-4. Semantic Representation of *The man that saw John saw Mary*.

All clauses, including subordinate clauses, are enclosed with double brackets. In order to signal subordination, the double brackets around subordinate clauses are slightly smaller than the double brackets that enclose independent clauses. If you look closely in Figure 2-4, you'll see that the brackets around the relative clause are slightly smaller than the brackets around the main clause.

Shown below in Figure 2-5 is a sentence that contains an object complement clause. Object complements are subordinate clauses, so they have the smaller double brackets, but they're not embedded in any type of phrase.

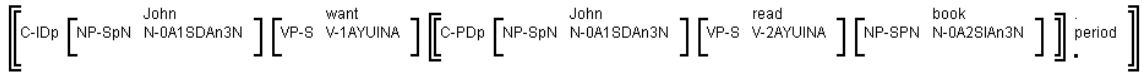


Figure 2-5. Semantic Representation of *John wanted to read a book*.

Figure 2-6 shows a sentence with an adverbial clause. All adverbial clauses begin with a preposition that signals the semantic relationship between the main clause and the adverbial clause.

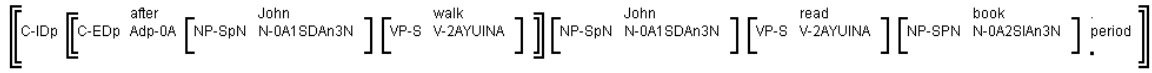


Figure 2-6. Semantic representation of *After John walked, he read a book*.

Figure 2-7 contains an attributive patient clause. Notice that it's a subordinate clause embedded in an adjective phrase.

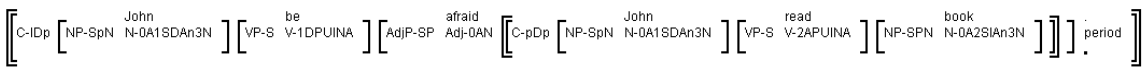


Figure 2-7. Semantic Representation of *John is afraid to read the book*.

You'll see many more examples of all these structures as you work through the tutorials.

2.4 Conclusion

You've now been introduced to the concepts, ontology, and semantic representations used in TBTA. These semantic representations contain a great deal of information; no language will use all of this information. As was mentioned above, the semantic representations reflect an English perspective. They use concepts that have generally been lexicalized by English, and the structures reflect English sentence structures. Subsequent tutorials will teach you how to restructure these semantic representations so that they become appropriate underlying representations for your target language. You'll be able to restructure the representations so that they contain words, features and structures that reflect your language rather than English. The remainder of these tutorials will teach you how to restructure the semantic representations and then synthesize the final target forms.