# ARRAU Annotation manual (TRAINS dialogues)

Ron Artstein and Massimo Poesio

June 23, 2006

# A general introduction to the task

The goal of this annotation is to collect data about the way context is constructed and used in natural language, particularly in relation to two phenomena normally called (nominal) *anaphora* and *deictic reference*.

The primary use of the term 'anaphora' is to describe what might also be called (linguistic) context dependence: namely, the fact that many expressions of natural language derive at least part of their meaning from the meaning of expressions previously produced in the same context. A clear example of context dependence are pronouns. The meaning of a pronoun is typically (although not always) related to that of a fuller linguistic expression already introduced in the linguistic context (e.g., produced as part of the previous utterances in the conversation). For example, in the fragment below, the pronoun it refers to the object already mentioned earlier in the same utterance using the phrase a boxcar. This object is called the antecedent of the pronoun.

M: I want you to take [a boxcar] from [Elmira] and load [it] with [oranges]

As most anaphoric expressions have a reduced form in comparison with the forms used to introduce entities in context, by using them language users achieve coherence (they connect the current utterance to previous ones) in a parsimonious fashion.

However, as said above, our primary interest in this study is not the (lexical) semantics of context-dependent expressions, but the way in which context is constructed and modified (an aspect of language often considered part of pragmatics). Therefore we are not only interested in cases in which the subsequent mention of an object is done using reduced forms; we want to track *all* mentions of an object through a text, including also cases in which these mentions are achieved using forms that normally would not be considered anaphoric, such as proper names like *Corning* or indefinite NPs like *a boxcar*.

(Anaphoric) context dependence can be more complex than in the example just seen. Some expressions derive their meaning from the linguistic context not in that they refer to an object already mentioned, but because they refer to entities which, while new in the discourse, are nevertheless tightly related to objects previously mentioned, in such a way that their reference would not be

understandable if such relation wasn't apparent. An example is shown below: the wheel refers to an object which hasn't been mentioned before in the dialogue, but it is nevertheless understandable because it is a part of an object – the boxcar at Elmira – mentioned in the previous sentence.

S: Bad news about [the boxcar at Elmira]. [They] tell me [the wheel] is broken and will have to be fixed

These expressions are usually called *bridging* references as the listener is required to 'bridge' the gap by identifying which relation holds between the expression and which previous expression. You will realize that a lot of expressions could be considered anaphoric in this sense; we only want to mark a few such cases however. These bridging references are not so common in the dialogues, but we will see more examples of them in the texts we will annotate later.

Language expressions may also depend on the visual, as opposed to verbal, context. An utterance like *Could you pass me the salt?*, uttered, e.g., at a restaurant, is usually understandable even if salt hasn't been mentioned before; the recipient can recover the referent of the expression *the salt* from the visual context. (An even clearer example is *Could you please close the window?*.) We will use the term *deictic references* to refer to these expressions. In the TRAINS dialogues, conversational participants often refer to objects in the map using deictic references. We want to mark these references, both because it's an easy way to check the consistency of the annotation, and to get data about context-dependent expressions whose interpretation can only be recovered from the visual context.

# Procedure to follow to annotate the TRAINS dialogues

- §1 In this annotation, your task is to mark *reference relations* in the dialogues you see on the screen, using the MMAX2 software, following the instructions below.
- §2 The dialogue is between two people who are trying to devise a plan for moving around rail cars in the railway network described in the map in Figure 1. You should consult the map as you read the dialogue.
- §3 Be sure to pay attention to the text while marking the dialogue; after you have completed the task you will be asked to write up a summary.
- §4 This manual contains both instructions on how to do the annotation (with examples) and technical instructions for using the software.

#### The MMAX2 software

§5 For the annotation you will use three windows (Figure 2): the dialogue window, the markables window, and the attribute window. Ignore all other windows that appear on your screen.

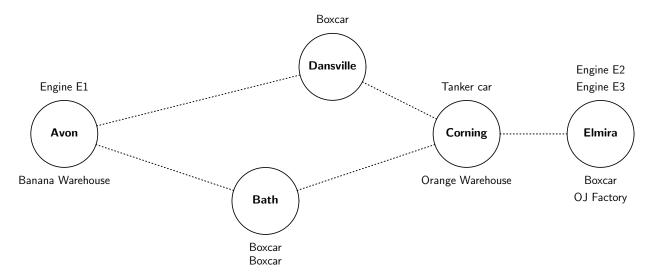


Figure 1: The TRAINS world

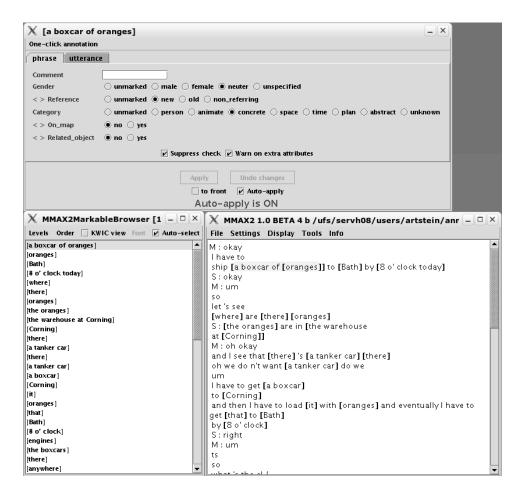


Figure 2: MMAX2 windows

- §6 Certain phrases in the text are designated as *markables*: they are delimited by brackets in the dialogue window, and also appear as a list in the markables window. *You select a markable by clicking on it with the left mouse button in the markables window*; the corresponding phrase is then highlighted in the dialogue window. You should go through the markables one at a time, clicking on each markable twice to select it before annotating its attributes as discussed below.
- §7 Each markable is associated with a series of *attributes*, and each attribute with a set of *values*. When a markable is selected, the corresponding attributes appear in the attribute window. Initially, the values of all attributes are set to "unmarked".
- §8 You should read the dialogue, and as you do so go through the markables in order and mark the appropriate attributes, as explained below.
- §9 The tool you are using has many features which are not described in this manual. If you do something accidentally and don't know how to get back, ask for help.
- §10 You must select a value for each attribute that appears in the attribute window. To do this, you first have to select the markable by clicking on it in the markables window with the left mouse button, and then click on the appropriate radio button in the attribute window with the left mouse button.

#### Comment

As you click on a markable which you haven't annotated yet, you will see three attributes: *Comment*, *Gender*, and *Reference*.

- §11 You should use the *Comment* attribute to indicate when you are unsure about the annotation of a markable, and possibly what you think the problem was and / or which alternatives you considered.
- §12 Note that the auto-apply feature does not work for the comment attribute. After entering your comment, you must click on the "apply" button in order to save it.

#### Gender

§13 The first thing you should do is to mark the (syntactic) gender of a markable using the attribute *Gender*. The possible values are "masculine", "feminine", "neuter", or "unspecified" when the NP could be either masculine of feminine (as in *the doctor*) or with coordinated NPs containing both masculine and feminine NPs (as in *John and Mary*).

#### Reference

Your next task is to annotate your interpretation of the markable. This begins by choosing a value for the attribute *Reference*.

#### **Term-denoting expressions**

Many types of natural language expressions can serve an anaphoric function, but in this study we are only interested in the anaphoric properties of **noun phrases** (NPs): expressions whose main word is a noun, like **the orange warehouse** or **a boxcar**, as well as proper names like **Bath** and pronoun like **it**. You will not have to find the noun phrases yourselves: they have already been identified and annotated as MMAX2's **markables**. (Some errors in identifying them – incorrect boundaries, missed noun phrases – will no doubt have taken place; it would be good if you could correct them.) You will however need to find which among these noun phrases are indeed anaphoric in the sense discussed above. One of your first tasks when analyzing a markable will be to identify what type of NP it is, using the attribute **reference**.

Anaphoric noun phrases are an instance of *term-denoting noun phrases*, so the first step towards identifying anaphoric expressions is to learn how to recognize these. A term-denoting NP is a noun phrase which is used to mention an object. The most typical examples of term-denoting NPs are proper names such as *Bath* or *Avon*, but many other types of noun phrases are term-denoting – including for example *two boxcars* in

There are [two boxcars] in Bath

#### Idioms

It is important however to realize that not all NPs are term-denoting. Perhaps the clearest case of non-denoting NPs are NPs occuring in *idioms* such as *what the heck*. In this idiom, the NP *the heck* does not refer to anything — i.e., the meaning of the idiom is not derived from the meaning of the expression *the heck*. Other idioms containing non-denoting NPs are *what a pain in the butt* (in which neither a pain nor the butt really refer to anything) or *kicked the bucket* (in which *the bucket* does not refer to any object in particular.

#### **Expletives**

A second fairly clear case of NPs that do not denote terms are certain uses of the pronominal expressions there and it. These words can be (and often are) used to refer to objects, as in the following two examples:

```
My brother finally bought a dog. [It] is a big grey Alsatian. (it = "the dog that my brother bought") To meet Prof. Rodgers, go to the NLP Lab. [He] is often [there] at this time of the day. (he = "Prof. Rodgers", there = "the NLP Lab")
```

However, in other cases, *it* and *there* only serve as 'placeholders' (in these cases, these words are called *expletives*). In the first example below, *it* does not refer to anything: it is only there because of the syntactic requirement that English finite clauses need to have a subject even when the underlying predicate does not have an argument to be filled by this subject. *there* in the second example is there for the same reason.

[It] takes an hour to get to Corning
[There] are two boxcars in Bath

#### **Predicates**

In the cases seen so far, the NP did not have a meaning at all. However, NPs may be non term-denoting even in some cases in which they do have a semantic meaning. For instance, in the example above, when we utter it is a big grey Alsatian, the NP a big grey Alsatian does have a meaning, but it is not used to introduce a second entity in the discourse, and it does not function as an argument of a predicate; rather, it is used to express a property of the dog that my brother bought. This is also the case for the NP the best chess-player in the school in Hillary is the best chess-player in the school. Again, the function of this NP is to ascribe a property to Hillary, not to introduce a new object in the discourse. In general, in most copular clauses (clauses whose main verb is to be, as in the two examples just discussed) either the subject or the object is a noun phrase used to express a property.

Another costruction in which NPs often serve as predicates are *appositions* – non-restrictive nominal modifications typically expressed using parentheticals. An example of apposition is the NP *my nephew* in

In this example, both the NPs my nephew and a terror express predicates, and therefore are not term denoting, whereas Carlo is term-denoting.

The test that you should use to decide whether an NP in one of the positions just discussed is predicative is to ask yourself whether the text is talking about two objects or just one. This test works reasonably well; however, you will find that in some cases it is difficult to identify which of the NPs in the clause is term-denoting and which instead is predicative. Both in the case of copular clauses and in the case of appositions you only need to consider one of the NPs as term-denoting. In order to decide which of these two NPs is term-denoting and which is a predicate, try to think which of the two could be viewed as a property of the other; but if you find you can't decide do not worry too much, just make sure you only include one in the coreference chain.

Apart from copular sentences and appositions, the NPs that are best viewed as expressing predicates or predicate modifiers are those that occur in locative expressions such as "to the left", or "go North", very common in some types of dialogue. "the left" and "North" are NPs, but in these examples the whole locative expression is best viewed as an adverbial not referring to any particular object. So you should not be concerned with marking such noun phrases as

deictic or anaphoric, the only exception being cases of anaphoric expressions explicitly referring to directions.

The interpretation of predicate NPs may also depend on context, but in this study we are not concerned with these dependencies. So you should not try to annotate anaphoric and deictic information (we will explain in a moment how this is to be done) about NPs you have identified as predicative.

#### Quantifiers

A fourth type of noun phrases that do not denote terms are *quantified* NPs. These include so-called *wh-NPs*, such as *which route* in

```
and then [which route] do you wanna take?
```

In this case it is fairly clear that the speaker is not referring to any object. Other examples of wh-NPs are where as in where is the engine now but also how long and how many oranges. Examples of quantifiers proper include all of the boxcars in

```
are [all of the boxcars] empty?
```

other quantifiers include expressions with the determiners every, as in everything; everything; everything; and everything; and everything; and everything; and everything; everything; and everything; ever

Quantified NPs are used to indicate the proportion of individuals of a set that has a certain property. They are never anaphoric in the sense we just discussed, so they should be marked in a special way using the *reference* attribute; however, the set itself (the *domain of quantification*) sometimes can be anaphoric. This domain of quantification can be indicated in two ways: either by means of a second NP, as in *all this / any of the boxcars* or by means of nouns, as in *any orange juice*. In the first case there should be two markables:

```
[all [this]]
[any of [the boxcars]]
```

so you can identify the domain of quantification as anaphoric, if necessary, by marking the markable embedded in the quantified NP. In the second case you should introduce a new markable:

```
[any [orange juice]]
```

#### How to annotate (non) term-denoting expressions

§14 The attribute *Reference* is used to specify whether a markable is term-denoting or not. This attribute takes one of the following values: "new", "old", or "non\_referring".

We will explain in this section how to choose the appropriate value. As you'll see, when you choose one of the values "new", "old", or "non\_referring" additional attributes appear in the window. We will explain now what to do in case you have chosen "non\_referring," and what to do in the other two cases in the next section.

§15 The values for the *Reference* attribute have the following interpretation:

**old:** A term-denoting markable in the sense discussed above, i.e., which refers to a concrete or abstract object which has already been mentioned or discussed earlier in the dialogue.

For example, in the following excerpt, the markable it refers to an object already mentioned using the phrase  $a\ boxcar$ .

M: I want you to take [a boxcar] from Elmira and load [it] with oranges

Examples of abstract objects are facts, events, actions, and plans. For instance, in the example just shown, M is proposing a plan: to take a boxcar from Elmira and loading it with oranges.

Note that the previous mention of the object need not have been made with a markable! (See discussion of *segment* below.)

We will explain below how you can indicate which previously mentioned object is being referred to in this case.

**new:** A term-denoting markable which in your opinion is the first mention in the dialogue of a concrete or abstract object. This value should also be used for markables that you want to mark as referring to an object associated with a previously mentioned object (see below).

non\_referring: A markable which does not refer to an object, whether concrete or abstract. If you choose this value, a new attribute, non ref type, will appear:

phrase utterance		
Comment		
Gender	○ unmarked ○ male ○ female ● neuter ○ unspecified	
< > Reference	◯ unmarked ◯ new ◯ old ◉ non_referring	
non_ref_type	○ unknown ● expletive ○ predicate ○ quantifier ○ idiom	
✓ Suppress check ✓ Warn on extra attributes		

and you will be requested to indicate why you think this markable is non-referring by choosing among the classes of non-referring markables discussed above.

(a) The markable is the word *there* or *it* used as placeholder. Choose the value "Expletive" if you think this is the case.

[There] are two boxcars in Bath

(= "two boxcars are in Bath")

[It] takes an hour to get to Corning

(= "to get to Corning takes an hour")

(Keep in mind that other uses of there and it do refer to concrete or abstract objects, and should be marked as "old" or "new" as appropriate!)

(b) The markable is a noun phrase used as a predicate, as discussed above:

The boxcar in Dansville is [a relic]

Use the value "Predicate" in this case.

(c) The markable is a noun phrase used as a quantifier, as discussed above.

```
Are there [any boxcars] in Dansville? Use the value "quantifier" in this case.
```

(d) The markable is part of an idiom, as in what [the heck]. Use the value "idiom" in this case.

Do not leave any markables with the value "unmarked" – its purpose is to serve as an indication that an appropriate value has not yet been selected.

### Marking anaphoric and deictic context-dependence

If you decide that a markable is "new" or "old," you'll be asked to provide more information about its reference. In this section you'll find instructions to do this.

**General criteria** As said above, our primary interest is the study of linguistic and deictic context-dependence. One of your most important tasks will be to identify cases of anaphora in the general sense discussed above: i.e., expressions referring to an object already mentioned. We already saw a few examples of expressions of this type, like the pronoun *it* in the following example:

```
M: I want you to take [a boxcar] from [Elmira] and load [it] with [oranges]
```

These cases should be marked by first choosing "Old" as the value of the "Reference" attribute, then proceeding as discussed below.

We also consider as anaphoric those expressions that refer to an abstract object such as a plan or an action. Such abstract objects may have been previously mentioned using a phrase (a markable), or using a segment — one or more utterances describing the plan, as in the following example, where this refers to the plan in the previous utterance by M:

```
M: Take [engine E1] to get [the boxcar] to Elmira. S: All right, [this] will take one hour
```

In addition, you will also be asked to mark deictic references to landmarks in the map. For example, in the following utterance, both *engine E2* and *Corning* are objects on the map shown earlier.

```
M: send [engine E2] off with [a boxcar] to [Corning] to pick up [oranges]
```

Some of these references refer to new objects, some to old objects. You should begin by choosing the appropriate value for the *Reference* attribute; in either case, a menu labelled *Object* will then appear, and you will then be able to choose the object referred to by the expression you are marking. (You'll find

that deictic references to towns using proper names, such as Corning, have already been automatically marked by us, but there are a few cases of references to towns - e.g., using pronoun there - that haven't been.)

In addition to subsequent mentions of objects already introduced in the discourse and deictic references to objects in the map you will also be requested to mark bridging references in the sense discussed above – references to objects which have not yet been mentioned, but are strictly related to objects which already have, as in the case of the wheel referring to a part of the boxcar at Elmira above. These cases are relatively rare in the TRAINS dialogues, but they are much more common in other types of text we will annotate later. In fact, they become so common that we will not be able to annotate all such cases; you should therefore in general concentrate on the following cases of bridging references:

part relations The wheel in the example above is an example of markable referring to an object that stands in a *part-of* relation to an object previously mentioned. You should mark these markables as *Related* using the instructions below.

set relations One class of relations of interest are those that hold between a set and its elements, or between a set and a subset. For example, after uttering there are two boxcars at Bath (thus introducing a set of two boxcars) a speaker may say we should send [one] to Avon or we should send [one boxcar] to Avon. Both [one] and [one boxcar] are references to elements of the set introduced in the previous utterance. Again, you should annotate these markables as Related. Or after saying We have [three engines] in total we may say [Two] are at Elmira; in this case, the markable [Two] refers to a subset of the set of three engines mentioned in the first utterance, and again should be marked as related.

**other relations** A third case of bridging reference we want to mark are expressions containing the word *other* and referring to a second object of the same type as an object already mentioned. E.g., in the TRAINS dialogues it is common to first talk about one engine, then introduce a second engine, and then return talking about the first engine by saying *the other engine*. These cases, as well, should be marked as **Related**.

misc You'll sometimes find clear cases of bridging references that do not fall into any of the categories above. You should set aside these cases for discussion and we'll decide what to do with them.

In most cases, the bridging reference refers to an object not previously mentioned in the dialogue. In these cases, you should first choose "New" as the value of *Reference*, then specify that the markable is "Related" to a previous object. After you've done this, you'll be able to identify the antecedent using a pointer.

In a few cases, you'll find that the bridging reference refers to an object which has already been mentioned. In these cases, you should always choose the value "Old" and identify the antecedent. However, you should also set the

**Related** value if the antecedent is further away than the object to which the markable is related.

#### Marking instructions: "new"

- §16 Just as in the case of "non\_referring", when you select "new" as the value for the *Reference* attribute, other attributes will appear.
- §17 The attribute (Semantic) *Category* should be used to mark the type of object the markable refers to, if any. Possible values include:

person: this value should be used for all markables that refer to people

(other) animate: use this value for markables that refer to other animate objects (animals, etc)

**(other) concrete:** to be used for all other concrete but not animate objects, such as engines, boxcars, etc.

space: for spatial locations (this includes towns, paths, or routes)

time: for all markables referring to temporal periods.

plan: for all markables referring to actions and plans.

(other) abstract: other abstract references which are not temporal expressions and are not plans (e.g., the law, injustice, art).

unknown: for markables that could be either abstract or concrete.

The *Category* attribute will also appear when you choose "old" as the value for *Reference*.

- §18 The next attribute, called <code>On\_map</code>, allows you to specify whether the markable refers to an object on the map. If you set this attribute to the value "yes", a second attribute will appear, called <code>Object</code>, which allows you to specify (using a menu) which object the markable refers to.
- §19 The next attribute, *related\_object*, is to be set to "yes" when the markable is a bridging reference in the sense discussed above. When you do so, you should also indicate the most recent mention of an object to which the reference of the markable is related, as follows.
  - (a) In the dialogue window, right-click (using the right mouse button) on the brackets that surround the phrase denoting the related object.
  - (b) The text "Mark this phrase as related" will appear; click on it with the left mouse button to set the mark.

This will cause a number to appear in the attribute window next to the **Related\_object** attribute, and will also cause an orange line to be drawn in the dialogue window between the markable and the phrase it refers to.

§20 NOTE: AT THE MOMENT WE HAVE TWO WAYS OF MARKING RE-LATIONS BETWEEN SETS AND THEIR ELEMENTS: THROUGH THE RELATED MECHANISM, AND THROUGH THE PLURALITY MECHANISM. INSTRUCTIONS BELOW WRITTEN UNDER THE ASSUMPTION THAT WE'LL USE 'RELATED' IN THE CASES DISCUSSED SO FAR - ELEMENT OR SUBSET OF SET FOLLOWS SET - AND USE PLURALITY TO MARK PLURAL REFERENCES WHEN ALL ANTECEDENTS ARE PRESENT. (BUT WE MAY WANT TO RETHINK AFTER A WHILE.)

Note: do not use the <code>Related\_object</code> attribute to mark the cases in which a plural anaphoric expression is used to refer to a set of objects introduced singularly before, as in <code>Kim saw Robin</code>. They had been <code>good friends at school</code>. In this case, mark they as a plural reference using the methods discussed below.

#### Marking instructions: "old"

- §21 If you select the value "old" further attributes will also appear, each on a separate line. You should use these attributes to specify which previously mentioned object the markable refers to.
- §22 The Category and On map attributes should be set as discussed above.
- §23 The next attribute is used to specify the *reference type* of the markable. The meanings of the values are:

**phrase:** Use this value if the markable refers to an object which was already mentioned using a markable. E.g.,

M: I want you to take [a boxcar] from [Elmira] and load [it] with [oranges]

**segment:** Use this value if the markable refers to an abstract object – for instance a plan, event, action, or fact – which was discussed in an earlier segment of the dialogue, but not referred to using a markable.

M: Take [engine E1] to get [the boxcar] to Elmira.

S: All right, [this] will take one hour

Note: subsequent references to the same abstract object should now be marked as "phrase", since the object discussed in the segment has now been mentioned using a markable.

- §24 If you select "phrase", then you should also indicate the most recent mention of the object that the markable refers to.
  - (a) In the dialogue window, right-click (using the right mouse button) on the brackets that surround the phrase which the current markable refers to.
  - (b) The text "Mark this phrase as antecedent" will appear; click on it with the left mouse button to set the mark.

This will cause a number to appear in the attribute window next to the "Single\_phrase\_antecedent" attribute, and will also cause a green line to be drawn in the dialogue window between the markable and the phrase it refers to.

it refers to the same object as a boxcar

this refers to the plan in the previous utterance by M

pick up [bananas] at Avon	pick up [bananas] at Avon
and Mark this phrase as antecedent	and
deliver [them] to Corning	deliver [them] to Corning

# §25 CHANGE INSTRUCTIONS DEPENDING ON HOW THE CODING SCHEME FOR BRIDGES IS DEFINED

In case the markable acts both as a mention of a previously mentioned object and as a bridging reference, and if the associated object is mentioned more recently, you should also specify the related object, as done above for the case of "new".

§26 If a markable refers to an object that was mentioned earlier in the dialogue more than once, mark a reference to the most recent mention.

M: I want you to take [a boxcar] from [Elmira] and load [it] with [oranges], then send [it] to [Corning]

the second *it* points back to the first *it* 

- §27 If you select "segment", then you should also indicate which text region the markable refers to. You do this by selecting the appropriate lines of text
  - (a) In the dialogue window, right-click (using the right mouse button) somewhere in a line that's included in the text region that the markable refers to.
  - (b) The text "Add this segment to antecedent list" will appear; click on it with the left mouse button to set the mark.
  - (c) Repeat this process until you have added all the desired lines.

This will cause a series of numbers to appear in the attribute window next to the "Segment\_antecedent" attribute, and will also cause a red line to be drawn in the dialogue window between the markable and each of the selected lines.

M: and while [it] 's [there] [it] should pick up [the tanker] [3sec] Add this segment to antecedent list	M: and while [it] 's [there] [it] should pick up [the tanker]
S: okay	S: akay
[3sec]	[3se <b>b</b> ]
and [that] can get	and [that] can get

# §28 ADD HERE INSTRUCTIONS FOR MARKING REFERENCES TO 'THIS' OR 'THE PLAN'

Cases that should not be annotated One of the main problems when doing anaphoric annotation is to avoid annotating too much. Almost all expressions could be interpreted as anaphoric in one way or the other, but not all cases of anaphoric reference are equally interesting; and in some cases, the referent is easy to identify automatically, or the existence of a semantic relation can be inferred from the text. Broadly speaking, there are three types of information that you should *not* annotate:

§29 As already mentioned, towns have been automatically marked. You need not worry about marking them, but they may serve as antecedents of anaphoric expressions.

- $\S 30$  Anaphoric references in which the anaphoric expression is one of the first or second person pronouns: I, you, and we. These are all term-denoting expressions, but in virtually all cases the referent can be recovered from other levels of annotation.
- §31 Semantic relations which are explicitly expressed by the text. These include the relations expressed by possessive constructions, such as the relation between E1 and its boxcar in E1's boxcar. (You will not find many examples of such constructions in the TRAINS dialogues, but there will be many in other texts we will annotate later.)

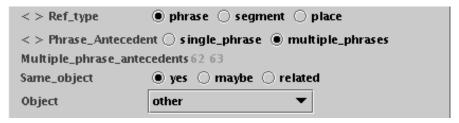
# Plurality and ambiguity

By default you can only mark one reference type, and if you choose "phrase" then you can only set a pointer to a single phrase or segment. Sometimes, however, you'll find that a markable may have more than one interpretation, or may refer to more than one object. These two situations should be marked in different ways, as follows.

§32 A *plural* markable is one which refers to a set of objects already mentioned using a phrase. For example, the markable *them* in the following snippet refers to the combination of the engine and the boxcar.

S: Please hook up [the engine] and [the boxcar], and send [them] to Elmira.

In these cases, when all elements of a set are introduced, a special mechanism to mark more than one pointer should be used instead of the *Related* mechanism discussed above. In the example above, two pointers from *them* should be marked – one to *the engine* and one to *the boxcar*. To set multiple *phrase* pointers, select the "multiple\_phrases" button.



Only mark multiple antecedents this way if they refer to distinct objects. If two possible antecedents refer to the same object, just mark the most recent one.

§33 By contrast, an *ambiguous* markable is one which has two (or more) *alternative* interpretations. For example, the markable *it* in the following snippet may refer either to the engine or to the boxcar, but not to both, although we can't determine which just on the basis of the text.

S: Be careful hooking up [the engine] to [the boxcar] because [it] is faulty

An ambiguous item should have **separate** markings for each interpretation: in the example above, it should have a "phrase" reference pointing to  $the\ engine$ , and a separate "phrase" reference pointing to  $the\ boxcar$ .

To specify a second reference, select the value "ambiguous" for the final attribute, *Ambiguity*. This will bring up a second set of phrase references, which you can use as before to mark the second meaning.

Notice that the alternative interpretations may be of the same reference type (phrase / segment) or of different reference types: i.e., a markable may be interpreted as referring either to a phrase or to a segment. You should mark each reference type separately.

- §34 A markable may also be ambiguous between an "old" and a "new" interpretation (i.e., whether it refers to an object already mentioned or to a new object). Markables may even be ambiguous between and "old" and a "non\_referring" interpretation. For example, in the following fragment it is not clear whether the *it* refers to a segment (the action of getting the boxcar to Elmira, say) or does not refer at all.
  - M: Let's take engine E1 to get the boxcar to Elmira.
  - S: I don't know how long [it] will take.

In order to mark these ambiguities, choose "old" as the value of the *reference* attribute, and then use the second reference type attribute to specify the values "new" or "non\_referring".

< > Ambiguity	🔘 unambiguous 🌘 ambiguous	
< > Ref_type_2	● phrase ○ segment ○ place ○ new ○ non_referring	
4 - Mariathean	Spinese Stegment Spinese Stien Stientering	
< > Phrase_Antecedent_2  single_phrase  multiple_phrases		
7 mac_mace. Single_pinac Sinakpa_pinacs		
Single_phrase_antecedent_24		

- §35 Only mark an item as "ambiguous" if the two interpretations are distinct, for instance if they refer to distinct objects, or if one interpretation refers to an object and the other does not. If two possible antecedents refer to the same object, just mark the most recent one.
- §36 If an item has more than two possible interpretations, only indicate the two most likely ones.
- $\S 37$  Sometimes, the ambiguity is temporary, but is resolved a few utterances later. For example, the pronoun it in utterance 3 might refer to the boxcar or the tanker, but utterance 5 makes it clear that the intention was the tanker.
  - 1. M: I want you to hook up the boxcar
  - 2. : and the tanker
  - 3. : and take [it] to Dansville
  - 4. S: Okay
  - 5. M: When you get there, fill [it] with Orange juice

In these cases, annotate the first markable as ambiguous. The second markable is, of course, unambiguous.

§38 In other cases, you'll find an 'ambiguity chain': a series of anaphoric expressions each 'ambiguous in the same way'. For instance, in the following fragment, the first it (in utterance 4.) may refer to either engine E1 or the boxcar; the same is true of the second it (in utterance 6).

1. M: good 2. : uhh

3. : can we please send engine E1 over to Dansville to pick up a boxcar

4. : and then send [it] right back to Avon

5. S: okay

6. : [it] 'll get back to Avon at 6

In this case, specify just one antecedent for the second it (that is, a pointer to the previous it), and use the value "ambiguous\_antecedent" for the *Ambiguity* attribute.

# **Correcting errors**

§39 If you make an error, you can correct it as follows:

to change the value of an attribute simply select a new value by clicking on the appropriate radio button in the attribute window; the old value will be removed.

to remove the reference of a phrase or segment markable follow the same procedure as for marking a reference, right-clicking on the brackets of the markable you want to remove; the text "Remove reference to this phrase/segment" will appear, and you can click that to remove the reference.

Be sure not to remove all the references of a phrase or segment markable! Also note that if you change the attribute value of a phrase or segment markable and then change it back, all the references you have marked will disappear and you will have to mark them again.