What is a Referent?

Referents are objects that are referred to in the text. Most of the time, referents are things that exist. Consider the following simple sentence, where the referents have been underlined:

\( (1) \quad \text{John kissed Mary.} \)

In this sentence, both referents are people – concrete things in the world. While this simple example covers a large number of cases, they are far from the full story. First off, referents may or may not have physical existence. In the next example the second referent is an abstract object:
(2) John had an idea.

Similarly, we can refer to things that don’t exist:

(3) If John had a car, it would be red.

The car does not exist, and yet we still refer to it. This sentence also illustrates an important point, namely, that a single referent can be mentioned several times in a text. In (3), it is the car that is mentioned twice. In this case, we say there is a single referent (the car), with two referring expressions (the phrases “a car” and “it”). These two referring expressions are called co-referential because they refer to the same referent. As in (3), we will use numeral subscripts to indicate that the two referring expressions co-refer.

An preliminary definition for referents is that referents are things that have been picked out for special attention. The trick, then, is to determine what ‘special attention’ means. In some sense, everything mentioned in a text, be it an object, an event, a time, a place, or something else, has been picked out for special attention, because it’s being talked about rather than something else of the same kind from the set of all possible things in the world. In (1), we might have marked “kissed” as a referent, since it is something happening, and we chose to talk about that rather than something else. But if everything mentioned is a referent, nearly everything in a text would be marked, which would be almost as uninformative as having nothing marked. What we are really interested in is reification, or, roughly speaking, items that find themselves the subjects or objects of verbs. To be more precise, if something is referred to using a noun phrase, it should be marked as a referent. Thus we have rule #1: Mark noun phrases as referring expressions.

This definition has the convenient property of having us to mark events (such as “kissed” above) only when they are picked out further beyond their use as a verb. Consider the sentence:

(4) John drove Mary to the office.

In (4) there are three noun phrases, and we do not mark the driving event as a referring expression, in accordance with our intuition. But if we appended a second sentence:

(5) John drove Mary to the office. It took forever.

We are picking out the act of driving as something interesting to talk about above and beyond its mere mention in the story, and in so doing we used a noun phrase “it” to refer to the event of driving. This forces us to “retroactively” (so to speak) mark the co-references to the event. Thus rule #2: Mark co-references of marked referring expressions, even if they would not normally be marked.

Pronouns and Numeric Expressions
Also mark referential pronouns as referring expressions, including possessive and reflexive pronouns.

(6) John was a doctor. He paid for his studies by himself.
In (6), different types of pronouns (possessive, reflexive) correspond to the same referent “John” and must be marked as co-references. Note that the phrase “his studies” contains two referring expressions: One to John’s studying (“his studies) and another to John himself (“his”).

Also make sure to mark numeric noun phrases as referring expressions. In (7) there are three noun phrases containing numerals.

(7) In the 1950s the city had 50,000 inhabitants. In 30 years the population doubled.

Keep in mind that numeric expressions which are not themselves noun phrases should not normally be marked, unless they co-refer with another referring expression:

(8) The furrow was fourteen feet high.

In this example, we do not mark “fourteen feet” as a referring expression. Rule #4: Pronouns and numeric noun phrases should be marked as referring expressions.

Generics

Noun phrases also might not refer to any object in particular. Take the following sentence:

(9) Lions are fierce.

Here we are not referring to a particular lion, but rather to a class, the set of all lions. These should be marked as referring expressions, but are different from particular lions:

(10) Lions are fierce. But Leo the Lion was the fiercest of all.

This indicates why it is important to mark generics. In (10), we would like to indicate what Leo the Lion was the fiercest of – namely, “all Lions.”

Despite this, we shouldn’t mark all generics, since almost everything is described as a member of some class of objects, e.g.,:

(11) Leo was a Lion.

Thus generics, like events, should be marked only when they are directly referred to – in other words, when the author intends to pick out the class itself, rather than merely indicating an object is a member of that class. Thus rule #4: Mark generics as referring expressions only when they are referred to directly.

Referential Extent: Modifiers

It is important to include in a referring expression not only the core noun or noun phrase that is doing the referring, but also to include any modifiers to the referring expression. This is because modifiers can substantially change the nature of the object being referred to. Compare the two following sentences:

(12) Every morning John woke early.
(13) **That morning John woke early.**

In (12), the noun phrase “every morning” refers to the set of all mornings, but (13) refers to a single morning with the phrase “that morning.” Similarly, you should include determiners (14), pronouns (15), adjectives (16), appositives (17), prepositional phrases (18), relative clauses (19), and other modifiers as part of the referring expression, as in the following examples.

(14) **The car was expensive.**
(15) **His car was expensive.**
(16) **The red car is expensive.**
(17) **The car, red as blood, was expensive.**
(18) **The car in the garage is expensive.**
(19) **The builder who erects very fine houses will make a large profit.**

Rule #5: **Quantifiers, determiners, pronouns, adjectives, appositives, adjectival phrases, relative clauses, and other modifiers should be included as part of referring expressions.**

Take into account that modifiers can themselves contain referring expressions. In example (20), where “Kent cigarette” is a modifier of the whole referring expression “Kent cigarette filters”, but at the same time a referring expression itself. The whole referring expression has been underlined, and the nested referring expression has been bracketed.

(20) **[Kent cigarette] filters contained asbestos.**

Therefore, in example (20) you will mark two referring expressions: “Kent cigarette” and “Kent cigarette filters.” Sometimes these rules lead you to mark rather large portions of text as referring expressions, with multiple nesting referring expressions (only the largest referring expression has been underlined; the rest are bracketed):

(21) **Takuma Yamamoto, vice president of [Fujitsu Motor]’s widgets and cogs division] since [June 1993], was fired yesterday.**

The underlined referring expression in (21) contains three internal referring expressions, namely, the car company, the car company’s division, and a date.

**Difficult Cases: Referring Expressions**

So far we have been considering some relatively straightforward referring expressions. Let us turn to a few more subtle cases, and techniques and conventions for handling them.

**Non-Referential ‘It’**

English has a device called a non-referential, or dummy, ‘it’. A non-referential *it* is used when there is no available argument to use with a verb (or the argument is already understood or can’t be spoken of directly), but the verb nevertheless syntactically requires an argument. In these cases we use a dummy it, and these should not be marked as referring expressions.
(22) It was raining.
(23) It was the fate of [the princess] to go to [the dragon].

**Negation**
Negation often creates conceptually tricky decisions. Noun phrases can express that no one thing is being referred to, as in (24), or referring expression may contain negations as modifiers that invert or otherwise alter their referent, as in (26). When a negation is used as a modifier to a referring expression, it should be included as any other modifier is included. A referring expression that refers to nothing or no one (or other empty set) should also be treated as a normal referring expression.

(24) No one is stronger than you.
(25) Nobody is stronger than you.
(26) He looked at nothing but himself.

Be careful in cases of verbs such as have and be, where the negation can be separated from the rest of the referring expression. In the following examples, a dotted line indicates words that are not part of the referring expression:

(27) I do not have any wine.
(28) He did not look at anything but himself.

**Conjunctions and Sets**
Things that have been previously referred to individually in a text are often later agglomerated into larger sets and those sets are then referred to directly. For example:

(29) Jack was a boy. Jill was a girl. They went up the hill.

In these cases, all the underlined referring expressions should be marked. In other cases, there is an implied set:

(30) [Jack] and [Jill] went up the hill.

All three referring expressions here should be marked: “Jack,” “Jill,” and “Jack and Jill.” This is because Jack and Jill are being referred to individually, and the set is used as an argument to the verb. Treat “or” the same as “and.” More complicated situations are as follows:

(31) [Jack], [Jill], and [Bill] went up the hill.
(32) [Jack] and [Jill] and [the Smith brothers] went up the hill.

In the first case we mark the set as well as the three individuals. We do not mark the sets (Jack,Jill), (Jack,Bill), or (Jill,Bill), because these are not syntactically picked out as separate things. On the other hand, in (32), the writer has gone out of his way to express the set in a way that is straightforwardly decomposable into the whole set, the Smith brothers, Jack, Jill, and “Jack and Jill.”
Articles and Possessive Pronouns
Be careful about the attachment of articles and possessive pronouns. Consider the following examples:

(33) The [dragon]'s lair
(34) [Her] [father] and [mother]

In these two cases, the article “the” and the pronoun “her” attach to the outermost referring expression. They do not, syntactically, attach to the inner referring expressions (“dragon,” “mother,” “father”). Keep a careful eye on where these modifiers attach is important, because modifiers can radically change the nature of the object being referred to.

‘Of’ Prepositional Phrases
Another class of referring expressions that can be tricky for determining co-reference are those of the form “X of Y”, e.g.:

(35) This has caused problems among a group of workers.

Does the phrase “a group of workers” contain one referring expression or two (one to the group of workers, and another to the set of all workers)? Consider these similar examples:

(36) Smoking has caused a high percentage of cancer deaths.
(37) Smoking has caused most cancer deaths.

One way of testing this is to try substituting the ‘Y’ for the ‘X’, and seeing if the fundamental class of the referent changes. If the class does not change, we have only a single referring expression. For example, in (35), we substitute “workers” for “group of workers”, we will still be talking about people. Thus we have only a single referring expression. In (36) the overall referring expression is to a percentage, but the internal object of the “of” prepositional phrases are “cancer deaths.” These are clearly different fundamental kinds of objects, and so there are two different referring expressions. By contrast, in (36), “most cancer deaths” is the same basic type as “cancer deaths”, and so we have only a single referring expression again.

Generics, or items that look like generics, also interact with of prepositional phrases in tricky ways:

(38) Will you not eat of my cake of rye?
(39) Have a cake of wheat.
(40) She came upon a river of [milk]. “Drink of my milk with [pudding],” said the river.

In these examples we do not mark “rye” or “wheat” because they do not refer to particular instances, but rather to general materials out of which the cakes are made. In (40), on the other hand, we do mark “milk” because it is later picked out in its own referring expression, and so we mark the other instances for co-reference purposes.
Existential vs. Locative ‘There’

Keep an eye out for the word “there” that is used in either an existential or locative sense. An existential use of there is shown in (41), and should not be marked as a referring expression, whereas a there that refers to a particular place is shown in (42).

(41) **There** once was a man from Nantucket.
(42) John Lennon sat **there**.

In some tricky cases it is not clear whether the there is existential or locative, such as:

(43) “Look, **there** is my stove!”

In these cases, it is up to the annotation team to discuss whether this is a locative or existential case, and mark it appropriately.

Neighboring or Nested Repeated References

Do not mark repeated references to the same object as separate referring expression. This includes cases where a modifying clause to a referring expression refers back to the referent itself, as with the word himself in (46).

(44) “**John, John, John.**” Mary said, shaking her head. “You are so naïve.”
(45) “**Mom, Mom, look what I found!**”
(46) **John, who himself was known to dislike spam**, refused the green eggs as well.

Difficult Cases: Co-reference

The rules elaborated above cover what to mark as a referring expression. Once you have determined that some set of tokens is a referring expression that should be marked, your second task to is to determine if the referring expression refers to a previously-introduced referent (it co-refers with already marked referring expressions), or if it introduces a new referent. We have already seen some unambiguous cases of co-reference. Let us consider more subtle cases.

Quantification

The first case is that of quantification:

(47) **Every day John woke early. One day he overslept.**

Are the two marked referring expressions here referring to the same day? The answer is no, as the phrase “every day” refers to a set of days (a fairly large set, in fact), and “one day” refers to a particular day. No problems here. But what about:

(48) **Every day the goose laid a golden egg. The woman could hardly wait for the egg.**

Are they the same egg? This is a bit trickier. It’s clear that there is more than one egg – in fact, one egg for every day. And it’s clear that the woman could hardly wait for each of them. But does “the golden
egg” refer to the set of all the eggs? One technique for determining co-reference is to vary the quantification of the second referring expression and see if it changes the meaning:

(49) Every day the goose laid a golden egg. One day, the woman could hardly wait for the egg.

In (49) it is clear the second referring expression is to a particular egg and is not co-referential with the first referring expression since the phrase “one day” breaks us out of talking about the things that happened “every day.” This indicates the proper way to look at (20): the phrase “every day” introduces a special context in which an object (the golden egg) is introduced and referenced. The context, in this case, does not continue into the next sentence, so in (20) we conclude that the two referring expressions do not refer to the same referent. (Note that this context effect is much like in (3) above, where we introduce an imaginary car in an alternate possible world.) This leads us to rule #6: with quantified referring expressions, use variation of quantifiers to test co-reference.

Plural Referring Expressions
Plural referring expressions can present some special problems for co-reference. Consider these cases:

(50) The three sons, stared at one another.
(51) Each of the sons, was strong but lazy.

Although both “at one another” and “each of the sons” are referring to each singular son, at the same time they are referring to all of them. So both referring expressions should be considered as co-references of “the three sons”. Thus remember that some quantifiers can produce plural referring expressions even though they are referring to a set of singular referents at the same time.

Copular Expressions
Determining co-reference can be tricky in copular (“X is a Y”) expressions:

(52) John, was a scientist.
(53) John, was the scientist.
(54) John, was not the scientist.

In (52) we know from the very syntax of the sentence that we are describing John as a generic scientist, and so we do not mark the phrase “a scientist.” In (53), we are describing John as a particular scientist (one perhaps we talked about earlier in the text), and so it is also co-referential. However, the introduction of “not” in (54) breaks the co-referentiality of the sentence, and we have referring expressions to two different things.

Synecdoche and Metonymy
Another common case is the use of synecdoche or metonymy, figures of speech in which a part of an object, or a closely related object, is used to refer to the whole object:

(55) The White House, announced a new economic stimulus plan today. The president and his staff, argued that previous efforts had fallen short.
In this case “The White House” is a closely-related object that is used to stand in for “The president and his staff.” Contrast, however:

(56) The owner of the orchard, often could be found pruning the old trees, and propping up the young ones.

In this case, “the old trees” and “the young trees” are not the same as “the orchard” – they are a part of the orchard, but not the same as it. The easiest way to discover this is to substitute one for the other, and determining if the sentence is (a) still well formed, and (b) the meaning remains unaltered. Thus rule #7: use the substitution test to determine appropriate co-reference relations.

Unknown Entities and WH-Words
Phrases or words whose actual referent is unknown at the time of reading should be marked normally as referring expressions:

(57) Whither did they go?
(58) Who here is a criminal?
(59) Prince Ivan set out to look for the woman he was to marry.

Difficulties arise when determining co-reference. It will be our practice to mark these referring expressions as co-referent with whatever referent is later determined to actually fill that role. Therefore:

(60) “Whither did they go,” she asked. “Thither!” he said.
(61) Prince Ivan set out to look for the woman he was to marry, ... Ivan took Maria to wife.

This will not be a satisfactory solution for stories or texts in which the final identity of the referent is unclear. If you come across these cases, bring them up to your annotation team.

Summary of Rules

<table>
<thead>
<tr>
<th>#</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mark noun phrases as references</td>
</tr>
<tr>
<td>2</td>
<td>Mark co-references of marked referring expressions, even if they would not normally be marked</td>
</tr>
<tr>
<td>3</td>
<td>Pronouns and numeric noun phrases should usually be marked as referring expressions</td>
</tr>
<tr>
<td>4</td>
<td>Mark generics as referring expressions only when they are referred to directly</td>
</tr>
<tr>
<td>5</td>
<td>Quantifiers, determiners, pronouns, adjectives, appositives, adjectival phrases, relative clauses, and other modifiers should be included as part of referring expressions</td>
</tr>
<tr>
<td>6</td>
<td>With quantified referring expressions, use variation of quantifiers to test co-reference</td>
</tr>
<tr>
<td>7</td>
<td>Use the substitution test to determine appropriate co-reference relations.</td>
</tr>
</tbody>
</table>

Glossary

**appositive** a noun or noun phrase that describes another noun or noun phrase directly adjacent.
**copular expression**  A sentence (or phrase) of the form “X v Y,” where X is the subject, Y the object, and v is a linking verb.

**copular predicate**  The object in a copular expression.

**co-referential**  The relationship that holds between two referring expressions when they refer to the same referent.

**linking verb**  Connects the subject being described with a description.

**metonymy**  see *synecdoche*.

**referring expression**  A set of words that indicates a referent. For every referent mentioned in a text there may be multiple referring expressions.

**referent**  Something that we talk about. Referents may be concrete or abstract, real or imagined; they may be objects, times, quantities, events, or any number of other things.

**synecdoche**  (a.k.a. metonymy) a figure of speech in which a part of an object, or a closely related object, is used to refer to the whole object.