Chapter 2

ANAPHORA

I shall not attempt to give a serious definition of anaphoric element, a task which presupposes an understanding of this aspect of language which is, in my opinion, not now available.
  — *Paul Martin Postal (1969:205)*

The term "anaphora", used several times above, will not be determined with any greater precision in this paper than is usual; and far from reducing the number of open questions about anaphora, I will actually add to that number.
  — *William C Watt (1973)*

2.1. What is anaphora?

ANAPHORA[1] is the device of making in discourse[2] an ABBREVIATED reference to some entity (or entities) in the expectation that the perceiver of the discourse will be able to disabbreviate the reference and thereby determine the identity of the entity. The reference is called an ANAPHOR,[3] and the entity to which it refers is its REFERENT or ANTECEDENT.[4] A reference and its referent are said to be COREFERENTIAL. The process of determining the referent of an anaphor is called RESOLUTION. By ABBREVIATED, I mean containing fewer bits of disambiguating information (in the sense of Shannon and Weaver 1949), rather than lexically or

---

[1]The terminology and many of the basic concepts described in this section are derived from Halliday and Hasan (1976).

[2]By a DISCOURSE we mean a section of text, either written or spoken, which is COHERENT in the sense that it forms a unified whole (Halliday and Hasan 1976). We do not restrict its length, nor do we limit the number of speakers in the conversation in the case of spoken discourse. For convenience, we will sometimes refer to the speaker and listener of a discourse, using these terms to subsume respectively the writer and reader of written text.

[3]This term is due to Edes (1968).

[4]Webber (1978a) distinguishes between a referent and an antecedent, calling "antecedent" the invoking description of which the referent is an instance — see section 5.4. We will not need to make this distinction, and will follow general usage, using the two terms interchangeably.

phonetically shorter (Hirst 1977a).[5] Note that one possible realization of an anaphor is as a complete void — an ellipsis; see section 2.3.13.

Two simple examples of anaphors are shown in (2-1) and (2-2):

(2-1)   Daryel carried a pewter centipede and a box to put <u>it</u> in.

(2-2)   Because Nadia was passing the sex shop, <u>she</u> was asked to buy half a kilo of pornography.

Here, *it* and *she* are anaphors with referents *a pewter centipede* and *Nadia*, respectively. In these particular cases, the referents occurred explicitly in the text and did so before the anaphor. Neither need be the case. In the next example, (2-2) is recast with the anaphor first:[6]

(2-3)   Because <u>she</u> was passing the sex shop, Nadia was asked to buy half a kilo of pornography.

That the referent need not be explicit is shown in these texts (the first based on an example of Grimes (1975:46), the second, Webber (1978a)):

(2-4)   When Ross visited his Aunt Cicely, <u>they</u>[1] spent the afternoon talking. Then, as arranged, Nadia arrived. Ross kissed his aunt goodbye, and set off with Nadia to the discotheque, where <u>they</u> [2] danced the night away.

(2-5)   Ross gave each girl a crayon. <u>They</u> used <u>them</u> to draw pictures of Daryel in the bath.

In (2-4), *they*[1] refers to the set {Ross, Aunt Cicely}, and *they*[2] to {Ross, Nadia}. Neither of these sets is mentioned explicitly, and the listener has to piece them together from the explicitly given elements. In particular, the MEANING of the text must be used to obtain the referent of *they*[2]. In (2-5), *they* and *them* are the sets of girls and crayons, respectively, whose existence is inferred from the first sentence.

> *There are no discounts on person-to-person calls. Check your phone book or the inside covers of this directory to see how and when these discounts apply in your area.*[7]

Conversely, an explicitly mentioned entity need not be referable — if negatively quantified, for example:

---

[5]Although most anaphors ARE lexically shorter than their antecedents, we shall later see some that are not.

[6]Strictly speaking, a reference which textually precedes its referent is called a CATAPHOR. Cataphors and anaphors are together called ENDOPHORS (see Halliday and Hasan 1976:14-18, 31-37). Again, we will usually be sloppy, and use the term *anaphor* to refer to both forms of endophor, except where repugnant to the context. Sometimes we will also include exophors (see below in this section).

[7]From an advertisement for the TransCanada Telephone System, 1978.

(2-6) Ross doesn't have a car. #It is a battered old Skoda.

(2-7) Ross doesn't have a car any more. It was completely destroyed in
an accident last week.

(2-8) Ross doesn't have a car, and if he did, it probably wouldn't run.

It is unacceptable to predicate anything of the non-existent car in (2-6), but
acceptable in (2-7) because the car's previous existence is implied.   In (2-8),
it refers not to the car Ross doesn't have, but to the one in the expansion of *did*
as *did have a car* that he might have.

Often, an anaphor with a non-explicit antecedent refers to something more
complex than a set of explicitly mentioned items. Consider these texts:

(2-9) The boy stood on the burning deck
      Picking his nose like mad.
      He rolled it[8] into little balls
      And threw it at his dad.[9]

(2-10) Ross sat in the corner, knitting madly. Suddenly he threw it down,
      and stormed out of the room.

What was thrown in each case is the PRODUCT of the previously described actions
and components, namely the results of the nose-picking and Ross's knitting,
respectively.

Sometimes the antecedent is nothing more than something brought to
mind by part of the text. Here are some examples:

(2-11) Ross wanted to NAIL the boards together, but Sue made him do it
      with TAPE

(2-12) Nadia dreams a lot, but seldom remembers them.

(2-13) When I first saw your gallery, I liked the ones of ladies.[10]

(2-14) Idi Amin is a bad joke, unless you are unfortunate enough to live
      there.[11]

(2-15) Early one morning at the end of August, a truck came up to the
      house. We loaded the paintings of the summer into the back, and
      closed and locked the doors. We stood on the porch and watched
      the truck drive off.
         "He is a careful driver," Jacob Kahn said. "I have used him

---

[8]This usage has been called the DESPICABLE *IT* (Corum 1973).

[9]From: Turner, Ian Alexander Hamilton. *Cinderella dressed in yella: Australian children's
playrhymes*. Melbourne: Heinemann Educational, 1969, page 104, rhyme 26116.

[10]From: Mitchell, Joni. The Gallery. On: Mitchell, Joni. *Clouds*. LP recording, Reprise RS6341.
The quoted text is the opening lines of this song; not all informants found it completely accept-
able.

[11]Not all informants found this sentence completely acceptable.

before."[12]

(2-16) Nadia wants to climb Mt Everest, and Ross wants to tour Africa, but neither of them will ϕ because they are both too poor.

(2-17) Ross and Nadia wanted to dance together, but Nadia's mother said she couldn't ϕ.

In (2-11) (due to Watt 1973:466) the referent of *do it* is clearly *fasten the boards together*, though this is only implied by the verb *nail*.[13] In (2-12) (which is due to Corum (1973)), *them* refers to Nadia's dreams. In (2-13), *ones* refers to the pictures brought to mind by the mention of the gallery. In (2-14) the referent *Uganda* for *there* is suggested by mention of Amin. Similarly, in (2-15), the arrival of the truck suggests the presence of the driver, and this is enough for him/her[14] to be referenced anaphorically. In (2-16) (from Webber 1978a), the elided verb phrase *do what she/he wants to do* is a single VP combining and abstracting its two antecedents *climb Mt Everest* and *tour Africa*, and in (2-17) (also from Webber 1978a), the ellipsis stands for *dance with Ross*.

EXOPHORS[15] refer deictically (Fillmore 1972) (that is, in a pointing manner) to items in the external world rather than in the text. For example, in (2-18):

(2-18) Pick <u>that</u> up and put it over <u>there</u>.

that and *there* are exophors whose referent in the real world is something that the situation, perhaps including physical pointing, makes clear to the perceiver of the text.

In summary, an anaphor is a reference whose antecedent is a concept or entity EVOKED implicitly or explicitly by the preceding text or situation.

2.2. Anaphors as references to entities in consciousness

In the previous section I described an anaphor as a reference that "the perceiver of the discourse will be able to disabbreviate". I now wish to elaborate

---

[12]From: Potok, Chaim. *My name is Asher Lev.* [1] Penguin, 1973, page 231. [2] Heinemann, 1972.

[13]Watt (1973) has called this phenomenon — verbs like *nail* which can have related concepts extracted from them as antecedents — PENETRABLE REEFS (cf Corum 1973).

[14]Most people sexistly assume the truck driver to be male, and hence find (i) jarring in the same context:

  (i) "<u>She</u> is a careful driver," Jacob Kahn said.

[15]The term *pragmatic anaphora* has been used for exophora by Hankamer and Sag (1976), and picked up by several other authors. The term is misleading, and will not be used here, as almost ALL anaphora is, in a sense, pragmatic (cf Morgan 1978; Partee 1978).

on this, and to qualify it.[16]

The qualification is to the words "will be able", which might better be "is expected by the speaker to be able". For when a speaker uses an anaphor, there is no iron-clad guarantee that the listener will in fact have the ability to resolve it. For example, the listener may have been busy thinking about something else and didn't even hear the referent of the anaphor; or, more frequently, the referent was mentioned so far back in the discourse that the listener has completely forgotten it, as (2-19) demonstrates:

> (2-19) Just as Carrie, played by Sissy Spacek, can be seen as another of De Palma's ambiguous women, as in *Obsession*, other parallels in the construction of the two films spring rapidly to mind. One can compare, for example, the extraordinary power of the final moments of the present film, in which the gentle, sunlit, Vaseline-lensed scene is shattered by a sudden horror that makes many people literally jump out of their seats, with that of *Obsession*, wherein the unexpected again happens, though this time in the negative sense that the expected does not happen.
>
> However, despite De Palma's skill, it is <u>her</u> acting that ultimately makes the film.

Here, few people, especially those not familiar with the films being discussed, would be able to resolve *her* as Sissy Spacek without consciously looking back through the text to find the referent. Anyone who didn't know that De Palma is male might have erroneously chosen him as the antecedent.

What is illustrated here is this: for an anaphor to be resolvable, its antecedent must be in what we shall for the time being call the listener's "CONSCIOUSNESS".[17] When a speaker uses an anaphor, they assume (usually correctly) that its antecedent is in the listener's consciousness and is therefore resolvable; if they are wrong, the discourse becomes ill-formed from the listener's point of view. Chafe (1970) has likened consciousness to a stage. Mentioning a concept, even implicitly, puts it on stage, from where it slowly retreats into the wings unless mentioned again. Concepts can be referenced anaphorically when and only when they are on stage (subject always to the constraints of syntax).

The speaker's assumption is apparently based on a model of the listener's consciousness which the speaker maintains (cf Winograd 1976). There have been no investigations into the nature of this model (but see Norman, Rumelhart and the LNR Research Group (1975:68ff) and Grosz (1977a, 1977b)), nor even has its psychological reality ever been shown. It is, however, probably part of a larger model of the listener that the speaker constructs, the necessity of which has been shown by Cohen and Perrault (1976), Perrault and Cohen (1977), and Cohen (1978), to mention but a few (cf also Webber (1978a)).

---

[16]The influence of Chafe (1972, 1974) and Nash-Webber and Reiter (1977) is evident in this section.

[17]For readability, I will not in future put the quote marks round *consciousness*. However, they should be understood as intended whenever I use the word. In section 3.2.1 I introduce better terminology.

2.2 Anaphors as references to entities in consciousness

How does an antecedent enter the listener's consciousness in the first place? There are four basic ways. The first, illustrated by examples (2-1) and (2-2), is that the antecedent be explicitly mentioned in the text, and further, as we have just seen, this mention must be "recent".[18]

The second is similar, except that the mentioning is implicit. We saw this in example (2-4), where things like set elements were given, causing the listener to be "conscious" of the set itself. Again "recency" is relevant.

The third and fourth ways antecedents may enter consciousness result in exophors when the entity is referenced. We saw the third illustrated by (2-18), a sentence which would be accompanied by pointing (or a similar gesture) to draw the listener's attention to what *that* is and where *there* is.

The fourth method is qualitatively different from the other three, in that the speaker does not deliberately cause the antecedent to enter the listener's consciousness. Instead, the speaker makes a calculated guess that other means have previously placed it there. Here is an example: The scene is a party at a wealthy person's home, and one of the guests is admiring a painting on the wall. The host comes up and says:

(2-20) Do you like it? It's an original Chagall.

The host can use *it* to refer to the painting because it is clearly the upper-most thing on the guest's mind at that moment — or at least so the host assumes. If in fact the guest was merely staring blankly into space in front of the painting, the guest would probably not realize at first what the host was talking about.

It follows from the above that if a computer is to take the part of a listener in discourse, it too must have a consciousness, or a model thereof, to understand anaphora. Further, if it is to generate discourse, it must make judgements about its listener's consciousness to use anaphora correctly; that is, it will need to model its hearer's consciousness. In this thesis we will be primarily concerned with the former case, namely modelling a listener's consciousness on a computer for anaphora resolution.

> *Anaone,*
> *Anatwo,*
> *Anathree,*
> *Anaphor!*
> — *Mark Scott Johnson*[19]

---

[18]Much of this thesis will be concerned with determining exactly what is meant here by *recent*.

[19]Personal communication.

2.3. Varieties of anaphora

Before you can resolve an anaphor, you have to know that it's an anaphor. This section, therefore, will be devoted to identifying the common or garden varieties of anaphora, and also a few more exotic species.[20]

2.3.1. Pronominal reference

The word *pronoun* has two meanings. Firstly, it can refer to a part of speech such as *he*, *she*, *it*, *they* or *that*. Secondly, it can refer to an anaphor whose antecedent is a noun phrase, that is one which "stands in place of a noun". In classical grammar, these meanings were generally taken to be equivalent. However, we shall see that they are not, and there are many cases in which pronouns in the first sense are not pronouns in the second sense, and vice versa. In this thesis, we shall generally use the word *pronoun* with its first meaning. To avoid confusion, we shall say that pronouns in the second sense of the word are PRONOMINALLY REFERENT.

Most pronouns ARE pronominally referent. For example:

(2-21) Ross bought {a radiometer | three kilograms of after-dinner mints} and gave {it | them} to Nadia for her birthday.

(2-22) Nadia wanted a gold ring, but Ross bought her a plastic one.

(2-23) Ross told Nadia about the coming of the Antichrist. It is due very soon, and he has bought exclusive film rights to it from the Vatican.

Pronouns are usually marked for gender and/or number, which is often useful in resolution. However, there are awkward exceptions. In this text, *she* refers to a person, film director Robert Bresson, who is probably marked as male in the listener's world knowledge:

(2-24) Who is this Bresson? Is she a woman?[21]

Similarly, in the novel *Even cowgirls get the blues*[22] the character named The Countess is introduced on page 63. It is not until page 66 that we find out that The Countess is male, and we are told this only implicitly by the author's referring to him by the pronoun *he* when there is no other possible referent. A human reader is momentarily fazed by this, but finds recovery easy.

---

[20]This section is an expansion of a similar section in Hirst (1976b). An alternative taxonomy appears in Nash-Webber (1977) and Webber (1978a).

[21]From: Robinson, David. Festival report: Berlin. *American Film*, *III*(1), October 1977, 68-70, page 68.

[22]Robbins, Tom. *Even cowgirls get the blues*. New York: Bantam, 1977.

A similar problem, which is becoming increasingly common, is the use of the SINGULAR EPICENE pronoun — a genderless plural third-person pronoun referring to a singular third-person of unknown, or deliberately unmarked, gender. For example:[23]

> (2-25) %The author thanks the reader for <u>their</u> kind indulgence.

> (2-26) %The most important qualification for the new programmer I want to hire is that <u>they</u> be fluent in Cobol.

> (2-27) "Would it not be possible for someone to come out by way of the drawing-room window and in this one while Mr Fitzroy was out of the room, and return the same way?" [asked Poirot.]
> %"But we'd have seen <u>them</u>," objected the Admiral.[24]

> (2-28) %Neither Ross nor Sue sank <u>their</u> teeth into my apple.

((2-28) is based on an example from Whitley (1978:19).) In many idiolects, these uses of *their*, *they* and *them* are acceptable substitutes for *his/her*, *he/she* (sometimes rendered as *s/he*) and *him/her*. Other idiolects fiercely reject such laxness in selectional restrictions, and such idiolects may be an unstated reason why some people virulently oppose current moves to "desex" language. A computer NLU system should be willing to give people the benefit of the doubt in this respect, and thus be able to understand text like the above examples, even though an occasional ambiguity may be thereby engendered.[25] For more discussion on the use of the singular epicene *they*, see Whitley (1978).

The horrible bureaucratese expression *same* acts like a pronoun with the special restriction that it can only refer to very recent noun phrases, usually the one immediately preceding it:

> (2-29) Persons using this coffee urn must clean <u>same</u> after use.

> (2-30) Complete the enclosed form and post <u>same</u> to the above address.

Interposing another noun phrase, *he/she* and *black ink* in the following examples, makes the sentence very marginal, at least in my idiolect:

> (2-31) ?When the user has finished with this coffee urn, he/she must clean <u>same</u>.

> (2-32) ?Complete the enclosed form in black ink and post <u>same</u> to the above address.

Intersentential reference with *same* also reduces acceptability:

---

[23]The symbol "%" indicates a sentence whose acceptability varies widely over different idiolects.

[24]From: Christie, Agatha. The submarine plans. in: *Poirot's early cases*, Fontana/Collins, 1974, page 130. [This text was originally published some time between 1923 and 1936.]

[25]The astute reader will have already noticed that this thesis is written in the lax idiolect.

(2-33) ?Complete the enclosed form. Post <u>same</u> to the above address.

2.3.2. Pronominal noun phrases: Surface count anaphors

Certain noun phrases also act as pronominal anaphors. These include *the former* and *the latter*. We shall call these SURFACE COUNT ANAPHORS:

- (2-34) Sue stared at the pumpkin and the turnip, and declared that she preferred <u>the former</u>.
- (2-35) One union, Prince Rupert Co-op Fisherman's Guild, owns a fish processing plant there. The other, the Amalgamated Shoreworkers and Clerks Union, represents workers in the plant. <u>The former</u> locked out <u>the latter</u> on June 23 when they couldn't agree on a contract for the workers.[26]

The former example suggests that ordinal numbers can also be construed as anaphors, as in (2-36):

- (2-36) Nadia removed from her bag a tissue, a dime and a crumpled dollar note, and absentmindedly handed the cashier <u>the first</u> instead of <u>the third</u>.

Although not great literary usage, this is syntactically correct and we understand its meaning. (See also Postal (1976).)

Surface count anaphors require not only that the antecedent be in consciousness, but also that the surface structure of the sentence (or at least the order of possible referents) be retained. This leads to the problem of what a possible referent for such an anaphor is. For example, (2-36) contains six noun phrases before *the first*; you probably didn't notice that there were so many, let alone count them. They are: *Nadia, her bag, a dirty tissue, a dime, a crumpled dollar note* and *the cashier*. (There is also an elided seventh, *Nadia*, before *absentmindedly*.) If *the first* and *the third* simply counted NPs in the sentence, their referents would be, respectively, *Nadia* and *a dirty tissue*, though we understand them unambiguously to be *a dirty tissue* and *a crumpled dollar note*. Clearly, to resolve such anaphors computationally, we need some way of knowing where to start counting.[27]

---

[26] From: Evans-Atkinson, Evan. From your side: Labor dispute causes waste of good fish. *The Vancouver sun*, 11 July 1978, page B6.

[27] One often sees sentences like (i) or, even worse, (ii) and (iii) in sloppy writing:

- (i) ?Ross was carrying a large box. <u>The latter</u> was brown.
- (ii) ??Ross entered the room with a box under his arm, and put <u>the latter</u> on the mantelpiece.
- (iii) ??We know well that potent insect Xylocopilpil, which is to the Xylocopid as the auk to

If there are too many items to be counted in a text with a surface count anaphor, the result is unacceptable, as not all possible referents can be retained in consciousness at once:

(2-37) On the twelfth day of Christmas my true love gave me eight ladies dancing, six drummers drumming, eleven songbirds singing, nine pipers piping, fifty lords a-leaping, seven federal agents, a swarm of swans a-swimming, five pogo sticks, four cauliflowers, three french fries, two cans of yeast and a parsnip in a pear tree. #I returned all but <u>the eleventh</u> to the store the following morning.

2.3.3. Pronominal noun phrases: Epithets

Epithets can also be used pronominally, as in these texts:

(2-38) Ross used his Bankcard so much, <u>the poor guy</u> had to declare bankruptcy.

(2-39) When John found out about Mary's marital infidelity, <u>the bastard</u> punched her.[28]

Lakoff (1976) has shown that epithets cannot have pronouns as their antecedents.[29]

2.3.4. Prosentential reference

Pronouns and words such as *such* and *so* may be PROSENTENTIALLY REFERENT. For example, consider (2-40) (due to Klappholz and Lockman 1975):

---

the hummingbird. <u>The latter creature</u> is but an inch overall. [From: Hepworth, John. Outsight: Shock! Horror! Giant bee stuffs Boeing 747. *Nation review*, *8*(32), 25-31 May 1978, page 20.]

The intention in (iii) is clearly that *the latter creature* is to refer to *the Xylocopid*, not *the hummingbird*. These texts are not acceptable in my dialect, though some people do not seem to mind (i) at least. For more of this, and its implications for transformational grammar, see Postal (1976).

[28]John and Mary are those playful characters well-loved by all readers of Schank (1975 and others).

[29]Apparent counterexamples to this can be explained as cataphora. For example, (i) parallels the structure of (ii) rather than (iii):

(i) When he entered the store, <u>the poor bastard</u> was robbed.
(ii) When he entered the store, <u>Daryel</u> was robbed.
(iii)When Daryel entered the store, he was robbed.

(2-40) The president was shot while riding in a motorcade down a major Dallas boulevard today; <u>it</u> caused a panic on Wall Street.[30]

Here, *it* does not refer to any of the preceding noun phrases, but to the whole situation of the president being shot while riding in a motorcade down a major Dallas boulevard today. In the next example (from Anderson 1976) *so* refers to a complete embedded sentence:

(2-41) Your wife was under the impression that you would be away tonight, and as you can see, I thought <u>so</u> too.

More than just a single sentence may be so referenced. For example, the first sentence of Chapter 11 of Tuchman's *A distant mirror* is (2-42):

(2-42) <u>Such</u> was the France to which Coucy returned in 1367.[31]

Such refers to the essence of all of Chapter 10.

2.3.5. Strained anaphora

Lakoff and Ross (1972) point out the frequent idiolectic acceptability of sentences like (2-43):

(2-43) John became a guitarist because he thought that <u>it</u> was a beautiful instrument.

The anaphor refers to the guitar, although this is only brought into consciousness by the noun phrase *guitarist*. Watt (1975) has called this phenomenon STRAINED ANAPHORA. Lakoff and Ross develop syntactic rules which explain why (2-43) is acceptable, but (2-44) and (2-45) are not:

(2-44) *The guitarist bought a new <u>one</u>.

(2-45) *John was a guitarist until he lost <u>it</u> on the subway.

In general, the antecedents of strained anaphors must be lexically similar to the actual words used in the text, such as *guitar* being similar to *guitarist*. Thus informants generally find (2-46) less acceptable than (2-43):

---

[30]Some instances of this type of sentence are idiolectically unacceptable to some people.

[31]Tuchman, Barbara Wertheim. *A distant mirror: The calamitous 14th century*. New York: Knopf, 21 September 1978, page 232.

(2-46) ?John became a flautist because he thought that <u>it</u> was a beautiful instrument.

Sentence (2-47) (due to Watt (1975:111)) is an apparent counterexample, in which the anaphor is not morphologically similar to its antecedent at all:

(2-47) The government's decision to annex Baja California as the fifty-fourth state was the second blow to freedom in <u>as many</u> weeks.

However, the lexical relationship seems to be enough for the anaphor to work like that of (2-43) (see also Watt 1973, 1975).

All this does not mean that such anaphors refer to the surface structure (or something just under the skin), and ignore semantics; for if this were the case, we could use the fact that *a ruler* can mean both a sovereign and a measuring stick to rewrite (2-48) as (2-49) (due to Carlson and Martin 1975):[32]

(2-48) The king picked up a measuring instrument and measured the lamp.

(2-49) *The ruler picked <u>one</u> up and measured the lamp.

Exactly what role semantics plays in this phenomenon is not clear. As Watt (1973) points out, the mere fact that *father* means *one who has sired a child*

---

[32]There are punning exceptions to this — a sort of non-elliptic syllepsis — varying in acceptability. Carlson and Martin (1975) offer (i) and (ii); the first is generally accepted, the second not:

   (i) Henry Block even looks like <u>one</u>.[33]
   (ii) *Frank Church has never been in <u>one</u>.

My explanation for the difference in acceptability is that the name must be sufficiently unusual for the hearer to notice its double meaning even before the punning anaphor is encountered in the text. Hence, we have:

   (iii)*Norman Smith is descended from <u>one</u>. [From which: a Norman or a smith?]
   (iv)*Kim Spencer wears <u>one</u>.
   (v) Nadia Talent is full of <u>it</u>.
   (vi)Tom Collins drinks lots of <u>them</u>.

Such puns really do turn up in real world text, as (vii) [from: *Time, 109*(22), 30 May 1977] shows:

   (vii)Not all the aliens are bad however. One who is not is Chewbacca (he <u>doesn</u>'t), the 8-ft.-tall wookie.

True (elliptic) syllepsis, as for example in (viii) (after Webber (1978a:105), who labels it zeugma),[34] involves a similar kind of resolution:

   (viii)Ross takes sugar in his coffee, ϕ pride in his appearance, and ϕ offence at the slightest innuendo.

Non-elliptic zeugma (that is, metaphor combined with syllepsis) probably don't exist in coherent English; elliptic zeugma is bad enough.

[33]It is also possible to interpret this text as meaning *Henry Block even looks like a Henry Block*, where a name like Henry Block is supposed to have associated with it a stereotyped image that a person with that name allegedly resembles:

   (i) I just met someone named Archie Bunker, and, by jove, he even looks like <u>one</u>.

[34]Authorities disagree on where syllepsis and zeugma differ from one another. I follow here the terminology of Fowler's *Modern English usage* (1968).

2.3.5 Strained anaphora

does not permit (2-51) (after Watt 1973:461) to be derived from (2-50):

> (2-50) Ross has sired a child, but none of his friends have seen it.

> (2-51) *Ross is a father, but none of his friends have seen it.

That is, *father* is an ANAPHORIC ISLAND (Postal 1969) in (2-50). Yet in the same paper, Watt offers this alarming example (1973:486):

> (2-52) Ross is already a father THREE TIMES OVER, but Clive hasn't even had ONE φ yet.

A fortiori:

> (2-53) Ross is already a father THREE TIMES OVER, but Sue hasn't even had ONE φ yet.

That is, adding contrastive stress can turn an anaphoric island into a penetrable reef. (See section 7.1 for a possible explanation for this).

2.3.6. Difficult indefinite uses of *one*

A phenomenon which at first seems to be related to strained anaphora is the indefinite *one*, as in this text:

> (2-54) Smoking gives one cancer.

This could be rephrased thus:

> (2-55) Smoking gives {a | the} {smoker | person who smokes} cancer.

This seems to parallel the case of (2-43) above. However, *things are not so simple*. Consider:

> (2-56) My boss makes one work hard.

> (2-57) Malcolm Fraser makes one sick.

> (2-58) Plutonium in the atmosphere makes one sick.

These mean, respectively:

> (2-59) My boss makes all those people he supervises work hard.

> (2-60) Malcolm Fraser makes everyone who is aware of him sick.

(2-61) Plutonium in the atmosphere makes everyone sick.

In each case, *one* means *all those whom <the subject of the sentence> affects* — again, an item implicitly placed in consciousness. This also holds for (2-54). Thus, we see that indefinite *one* is not a particular case of strained anaphora.

> *Have you seen my wife, Mr Jones?*
> *Do you know what it's like on the outside?*
> — *Robin and Barry Gibb*[35]

2.3.7. Non-referential pronouns

Some instances of the pronoun *it* don't refer to anything, and hence are not anaphors, and some have referents defined by convention which need not be present in consciousness. It is necessary to recognize all these when they are found, lest precious hours be lost in bootless searches for textual referents.

Consider (2-62):

(2-62) It is fortunate that Nadia will never read this thesis.

This is a simple case of a dummy subject in a cleft sentence, derived from (2-63), and the *it* is essentially meaningless:[36]

(2-63) That Nadia will never read this thesis is fortunate.

Note that syntax alone is not enough to identify the dummy *it*. In (2-64) *it* is a dummy subject, but in (2-65) it could refer to *the cat*.

(2-64) This thesis contains many facts that would embarrass Nadia if she knew they were being published. It is therefore fortunate that Nadia will never read this thesis.

(2-65) If Nadia were to read this thesis, she would probably get so mad that she would kick the cat. It is therefore fortunate that Nadia will never read this thesis.

However, cleft interpretation seems to be the default in (2-65).

Some pronouns have conventional unspecified referents, as in this:

---

[35]From: New York mining disaster, 1941. On: Bee Gees. *Best of Bee Gees*. LP recording, Polydor 5837063.

[36]One could say, for convenience, that it does have a referent, namely *that Nadia will never read this thesis*, but this is merely playing with the definition of *referent*. There is, notwithstanding this, a clear qualitiative difference between this and other uses of the word *it*.

(2-66) It is half past two.

This could be restated thus:

(2-67) The time is half past two.

But the same process cannot, of course, be applied to (2-68) to give (2-69):

(2-68) It is half a lamington.

(2-69) *The time is half a lamington.[37]

In general, we have to be on the lookout for cases where *it* means by convention *the time*. Care is required, as we see here:

(2-70) How late is it? — It's ten to one.

(2-71) What's the starting price of Pervert's Delight? — It's ten to one.

There are other awkward cases, too.[38]

---

[37]This sentence is unacceptable for selectional or semantic reasons, while it is syntax that prevents (i) from being optionally rendered as (ii):

    (i) What time is it.

    (ii) *What time is the time?

[38]The question of these uses of *it* is complex and the only important point here is that they must be recognized by an anaphor resolver to avoid wasting time on fruitless searches for their referents.

Some of the problems in this area can be seen by considering (i):

    (i) It is raining.

Unlike the case of *the time*, we cannot simply eliminate this sentence's *it* by rephrasing:

    (ii) *{The sky | The weather} is raining.

But note also that (iii) is an acceptable sentence, although (iv) is strange to most people:

    (iii) It was half past two and raining when Sue finally arrived.

    (iv)?It was raining and half past two when Sue finally arrived.

Sentence (iii) suggests that *it* can mean both the time and the weather taken together — perhaps the general state of affairs. The strangeness of (iv) then needs to be explained. I leave this as an exercise for the reader.

The question of how and why *it* actually appears in these sentences is a matter of much debate in linguistics. Sentences like (v) (due to Morgan 1968) are even harder than (i) to rephrase without it:

    (v) It is dark outside.

However, such sentences may have non-dummy subjects in other languages, indicating the presence of a subject in a deep, language-independent representation of the sentence. For example, in German, the dummy-subject sentence (vi) translates into English as (vii) with a substantial subject:

    (vi)Es klingelt. [Literally, "It rings".]

    (vii)Someone is ringing.

See Morgan (1968) for a slightly different approach to this question.

2.3.7 Non-referential pronouns

2.3.8. Pro-verbs

The only English pro-verbs are forms of *to do* as in (2-72) and (2-73):

> (2-72) Daryel thinks like I <u>do</u>.

> (2-73) When Ross orders sweet and sour fried short soup, Nadia <u>does</u> too.

The antecedents are, respectively, the VPs[39] *thinks* and *orders sweet and sour fried short soup*.

Under certain conditions the antecedent can be two or more VPs. Nash-Webber and Sag (1978) cite this example:

> (2-74) She walks and she chews gum. Jerry <u>does</u> too, but not at the same time.

Of course, not all occurrences of *to do* are anaphoric: it can also mean *to perform <some task>*, and it can be a meaningless auxiliary:

> (2-75) Nadia <u>did</u> her exercises.

> (2-76) Ross <u>does</u> not like lychees with ice cream.

2.3.9. Proactions

When used in conjunction with *so, it* or demonstratives, *do* can reference ACTIONS in a manner which is almost prosentential. Consider:

> (2-77) Daryel frequently goes to the cupboard, where he secretly pours himself a glass of Cointreau. He drinks it in one gulp. Sue <u>does it</u> too, but less discreetly.

---

[39]Halliday and Hasan (1976:114-115) give examples in which *do* replaces only part of a verb phrase:

   (i) Does Granny look after you every day? — She can't <u>do</u> at weekends, as she has to go to her own house.

   (ii) Mrs Birling: I don't understand you, Inspector.
Inspector: You mean you don't choose to <u>do</u>, Mrs Birling. [From: Priestly, J B. *An inspector calls.* in: *The plays of J B Priestly.* Heinemann, volume 3.]

   (iii)What are you doing here? — We're mycologists, and we're looking for edible mushrooms. — Yes, we are <u>doing</u> too.

However, this usage is acceptable only in a British dialect of English; informants who were speakers of Canadian, American or Australian English immediately marked such sentences as British, and said that their dialect would not generate them. These dialects would use an ellipsis instead of *do*.

(2-78) Ross makes his dinner on weekdays, but when she stays the week-
end Sue <u>does it</u> for him.

(2-79) Nadia removed a herring from her pocket and began to fillet it.
Ross <u>did so</u> too.

In each of these texts, the PROACTIONAL anaphor refers not to the previous
events but to the action therein: to the act of taking a herring from a pocket
and beginning to fillet it, rather than Nadia's specific performance of that act.
Note in particular that (2-79) does not mean that Ross removed the herring
from Nadia's pocket, but rather from his own, and in (2-77), Sue pours herself,
not Daryel, a glass of Cointreau. However in (2-78) Sue cooks Ross's dinner, not
her own.

There is no firm dividing line between proactions and pro-verbs: (2-73)
could have *does it* or *does so* in place of *does* without changing its meaning.

2.3.10. Proadjectives

Postal (1969:205) points out that words like *such* are anaphoric in texts like
these:

(2-80) I was looking for a purple wombat, but I couldn't find <u>such</u> a wom-
bat.

(2-81) I was looking for a wombat which spoke English, but I couldn't find
<u>such</u> a wombat.

Such references are PROADJECTIVAL, or, in Postal's term, PRORELATIVE, referring
here to *purple [wombat]* and *[wombat] which spoke English*. Often the
antecedent is only implicit, as in (2-82):

(2-82) Ross came rocketing out the door and tripped over Nadia's nar-
balek, which bounced off and cowered under the garage. <u>Such</u> situa-
tions have been a common occurrence since the vacation started.

Here, the antecedent for *such [situations]* is not *[situations] in which Ross
comes rocketing . . .* but rather something like *chaotic [situations]*. See also
Halliday and Hasan (1976:76-87).[40]

---

[40]In bureaucratese and legalese, *said* can be used as a proadjective for very explicit discourse
cohesion:

(i) I bequeath absolutely my bandicoot Herbert to Ross Frederick Andrews of 79 Lowanna
Street Braddon in the Australian Capital Territory provided that the <u>said</u> Ross Frederick
Andrews shall keep feed and generally maintain the <u>said</u> bandicoot in good health order
and condition.

The said Ross Frederick Andrews means *Ross Frederick Andrews of 79 Lowanna Street Brad-
don in the Australian Capital Territory*. The *saids* serve to explicitly prevent the condition be-

2.3.11. Temporal references

The word *then* can be used as a anaphoric reference to a time or an event, as can *at that time*:

(2-83) In the mid-sixties, free love was rampant across campus. It was then that Sue turned to Scientology.

(2-84) In the mid-sixties, free love was rampant across campus. At that time, however, bisexuality had not come into vogue.

Many temporal relations such as *afterwards* are anaphoric, in the sense that the time they are relating to is also a referent determined like that of an anaphor. In (2-85), *many years later* implies a reference to *the mid-sixties*, in a very similar manner (though of course with different meaning) to the *then* of (2-83):

(2-85) In the mid-sixties, free love was rampant across campus. Many years later Sue turned to Scientology.

2.3.12. Locative references

The word *there* is often an anaphoric reference to a place:

(2-86) The Church of Scientology met in a secret room behind the local Colonel Sanders' chicken stand. Sue had her first dianetic experience there.

Locative relations, like temporal relations, may also reference anaphorically:

(2-87) The Church of Scientology met in a secret room behind the local Colonel Sanders' chicken stand. Across the street was a McDonald's where the Bokononists and The Church Of God The Utterly Indifferent had their meetings.

---

ing satisfied by a different Ross Frederick Andrews, or by the maintenance of a different bandicoot.

2.3.13. Ellipsis: The ultimate anaphor!

Some anaphors are completely null. In (2-88):

(2-88) Ross took Nadia and Sue ϕ Daryel.

the word *took* has been elided. A whole VP may be elided:

(2-89) Nadia brought the food for the picnic, and Daryel ϕ the wine.

Here the elided VP is *brought to the picnic*. VP ellipsis cannot in general be exophoric (see Hankamer 1978; cf Schachter 1977 and Hankamer and Sag 1976).

The above examples illustrated VP ellipsis. However almost any part of a sentence can be elided:

(2-90) Ross carefully folded his trousers and ϕ climbed into bed.

(2-91) Who put this bewildered baby bandicoot in Barbara's biscuit barrel?
— Ross ϕ.

In (2-90), the subject NP *Ross* is elided, and in (2-91) only the subject NP remains after the removal of *put that bewildered baby bandicoot in Barbara's biscuit barrel*. This latter kind of ellipsis is very common in answers to questions, so it is important that it be understood by any system which accepts natural language answers to queries.[41]

2.3.14. An awkward miscellany

The following examples are awkward exceptions to normal pronominalization:

(2-92) Andy sends the *1978 World Book Science Year Annual* to Lorri Dunn, 12, of Visalia, Calif., for her question:

---

[41]It should be noted that not all "syntactic gaps" are anaphoric. Thomas (1979) distinguishes ELISION and NON-REALIZATION, which are non-anaphoric, from true ellipsis, which requires context for its resolution. Elision is the removal of certain words, usually in informal speech, that may be recovered by applying certain conventional *rules of* conversation which Thomas details. An example:

(i) ϕ Got the tickets?

Non-realization is the syntactic removal, at a level below the surface, of elements that do not require recovery at all. An example of this is the non-appearance of [*by*] *someone* when (ii) is passivized to become (iii):

(ii) Someone murdered Jones.
(iii)Jones was murdered.

Why is <u>it</u> called a gunny sack?[42]

(2-93) Nadia: Is <u>it</u> pronounced "tom-AY-to" or "tom-AH-to"?
    Ross: Is WHAT pronounced "tom-AY-to" or "tom-AH-to"?[43]

One could dismiss (2-92) as an illiteracy — it is unacceptable in my idiolect — but (2-93) is quite acceptable.[44] Another version of (2-92) is also mysterious:

(2-94) Why is a gunny sack <u>so-called</u>?

Here *so-called* is an adjective which refers to a noun phrase — a most unusual state of affairs. It may be objected that the referent here is the adjective *gunny*, not the NP *gunny sack*. But consider:

(2-95) Why is psittacosis <u>so-called</u>?

(2-96) Why is rappelling <u>so-called</u>?

Here the referents are unquestionably NPs. Note that in (2-96) the NP is a gerund; this seems to be the only way to ask such a question about a verb.

---

[42]From: Andy. Ask Andy. *The province*, 11 July 1978, page 14.

[43]Old joke, recently resurrected on the television series *The muppet show*.

[44]That Nadia's question in (2-93) is well-formed is shown by Ross's reply being humorous. The humour relies on Nadia's question being quite acceptable, although based on a prototype that normally wouldn't be. See Hirst (1979) for more discussion.

2.3.14 An awkward miscellany

2.3.15. Summary of anaphors

The following table summarizes the different types of anaphor we have seen:

Type of anaphor	Lexical Realization
Pronominal	
• pronouns	*he, she, it, they, one,* . . .
• epithets	*the idiot, that stinking lump of camel excrement,* . . .
• surface count	*the former, the latter, same,* low ordinals, . . .
Prosentential	*it, so,* . . .
Pro-verbial	*do*
Proactional	*do so, do it*
Proadjectival/prorelative	*such, so,* . . .
Temporal	*then,* temporal relations
Locative	*there,* locative relations
Ellipsis	ϕ

2.4. Where does anaphora end?

The previous section dealt with various anaphoric proforms. The spirit of anaphora is not limited to proforms, however. This section examines some other linguistic constructions that can be used in an anaphor-like manner.

2.4.1. Paraphrase

Paraphrase is a restatement of a part of a text in different words to clarify the intended meaning or for stylistic reasons. When a noun phrase is subsequently paraphrased in a text, the result is often anaphor-like. Indeed, it is not clear where anaphora ends and paraphrase begins. Consider these examples:

(2-97) The man carrying the aeolian harp stumbled and for a moment Sue thought <u>the man</u> would fall.

(2-98) Sue watched the man from her hiding place. <u>The man</u> had an aeolian harp, which he was holding above his head in an attempt to make it play.

(2-99) At first Ross couldn't locate the Pope. Then he looked up, and saw <u>the beloved pontiff</u> floating gently to earth.

In (2-97), *the man* (second occurrence) refers to the man carrying the aeolian harp. Such INCOMPLETE REPETITIONS clearly fit our definition of an anaphor, although people may not always classify them as such. The problem of understanding them differs from the case where a proform is used only in the quantity of information given in the reference. Programs such as Bobrow's (1964) STUDENT (see section 3.1.1) have dealt with such incompletes, using heuristics to equate them with their referent. Further, as in (2-98), a single complete repetition is again anaphor-like in the way it performs a subsequent reference to the man with the aeolian harp.

In (2-99), *the beloved pontiff* refers to the Pope. Although this is not an abbreviation,[45] but rather a disabbreviation, it again shares the spirit of anaphora, and again the problem of understanding and making the connection is similar.[46]

The style of writing in which the paraphrases are not just lexically longer but are used to give more information than the original noun phrase occurs frequently in North American newspaper reports; (2-100) demonstrates this style:

(2-100)   BIG BEN FATIGUED

LONDON — With a rattle and a bang, London's famous landmark, the Big Ben clock, ground to a halt today at 4:46 a.m.
      <u>The 117-year-old timepiece</u> apparently was the victim of metal fatigue.[47]

Here the paraphrase (underlined) gives us new information, in this case about the age of the clock. We can make the connection easily since *the . . . timepiece* clearly points back to *the . . . clock*. (If the noun phrase had been *a . . . timepiece*, then the indefinite article would mean that a different clock was being talked about.)

In the next example, there is no definite article or other pointer to help resolve the coreference:

---

[45]It is not an epithet either, as it can be stressed if spoken.

[46]Could we take this analysis backwards, and construe *the Pope* as a cataphor of *the beloved pontiff* as we did in footnote 29? We probably cannot since, without more context, we could replace the latter but not the former with the anaphor *him*. In other words, in the absence of a compelling reason to do so we are loathe to allow the possibility of a cataphoric noun phrase existing where a cataphoric pronoun could not.

[47] Associated Press, 5 August 1976.

(2-101)   CHOWCHILLA, Calif. — Two men charged with the abduction of 26
school children appeared in a packed courtroom today amid tight
security and pleaded not guilty to 43 charges of kidnap and robbery.
James Schoenfeld and Frederick Woods both 24, appeared in
justice court with Schoenfeld's brother, Richard, 22, who entered a
plea of not guilty to the same charges a week ago.[48]

The two paragraphs of (2-101) could be two separate court report summaries;
only our knowledge of the style (and perhaps previous knowledge of the
Chowchilla kidnapping case) allows us to detect that *James Schoenfeld and
Frederick Woods* are the *two men* of the previous paragraph, and *justice court*
is *a packed courtroom*.

It is necessary, however, that the identity of the paraphrase and its
referent be reasonably easy to infer. Informants frequently failed to recognize
the paraphrase in this text:

(2-102)   Most of the city's federal buildings were dark, but chandeliers shone
brightly from the National Portrait Gallery. Inside the building in
which Walt Whitman once read his poetry to wounded Union troops
and Abe Lincoln held his second Inaugural Ball, a black-tie assem-
blage of guests stood chatting.[49]

In fact, *the building in which Walt Whitman once read his poetry to wounded
Union troops and Abe Lincoln held his second Inaugural Ball* is the previously-
mentioned National Portrait Gallery, but many readers assume two separate
buildings are being spoken of, apparently due to the difficulty of detecting the
paraphrase in such convoluted prose.

Not only NPs but also sentences and situations may be paraphrased. In this
example (after Phillips 1975) *the mistake* refers prosententially to the whole
preceding sentence:

(2-103)   Ross put his car into reverse instead of drive and hit a wall.
The mistake cost him two hundred dollars.

2.4.2. Definite reference

The anaphora and paraphrase problems are actually special cases of the
definite reference problem. This is illustrated in the next two examples:

---

[48] Associated Press, 4 August 1976.

[49] From: Davidson, Ralph P. A letter from the publisher. *Time*, *111*(20), 15 May 1978.

(2-104)   Nadia bought a DECsystem-10. <u>The processor</u> is a KL10B.

The scene for the second example is similar to that for (2-20), except that this time the guest is admiring the host's new car. The host comes up and remarks:

(2-105)   Because I'm a nostalgic horse racing fan, I've had <u>the speedometer</u> marked in furlongs per hour.

In these examples, the NPs *the processor* and *the speedometer* mean those of the DECsystem-10 and the car,[50] respectively, and semantically stand in the relation PART OF to those antecedents. Other possible relations include SUBSET OF (Klappholz and Lockman 1975), and ASPECT or ATTRIBUTE OF.[51] We see that anaphora and paraphrase are merely cases of coreferentiality where the relation is IS IDENTICAL TO.

Sometimes a coreference relationship is not one of those just mentioned, but rather is one determined by inference (Clark 1975). Consider this example:

(2-106)   "It's nice having dinner with candles, but there's something funny about the two we've got tonight", Carol said. "They were the same length when you first lit them. Look at them now."
John chuckled. "<u>The girl</u> did say one would burn for four hours and the other for five", he replied. "Now one is twice as long as the other."
They had been burning for the same time, of course. How long was that?[52]

The relationship between *the candle* and *the girl* is that the latter presumably is the salesperson who sold John the former. To determine this requires a high level of inference, such as that performed in the MARGIE system (Schank, Goldman, Rieger and Riesbeck 1975; Rieger 1975), and we would not want to say that there is an intrinsic semantic relation between girls and candles. A simplistic resolution algorithm would probably have decided that *the girl* in this example

[50]We regard *the speedometer* as a reference to *the car* with the latter as antecedent, rather than a direct reference to *the speedometer* as an item in consciousness, on the reasonable assumption that the speedometer itself was not in the listener's consciousness. Clearly, the speaker could have referenced ANY part of the car from the engine through to the little switch that makes the light come on when you open the door but it is unlikely that the listener would have had all these parts in consciousness.

[51]Examples of these relations:
SUBSET OF:
(i) The Department has graduated five students this year. <u>The PhDs</u> were all in AI.
ASPECT or ATTRIBUTE OF:
(ii) For Christmas that year, Julian gave Sissy a miniature Tyrolean village. <u>The craftsmanship</u> was remarkable. [From: Robbins, Tom. *Even cowgirls get the blues.* New York: Bantam, 1977, page 191.]

Klappholz and Lockman (1975) suggest MEMBER OF as another possible relation, but I am not convinced that it differs in practice from PART OF.

[52]From: Hunter, J A H. Figure it out. *The Canberra times*, 26 October 1977, page 25.

2.4.2 Definite reference

was *Carol*.

Between the extremes of a fixed relation like ASPECT OF and inferred relation like that in (2-106) is the vague relation CLOSELY ASSOCIATED WITH:

> (2-107)   The manager ushered Sue and Nadia into his <u>office</u> with obvious embarrassment.

The concept of *office* is closely associated with the concept of *manager*, through some fairly direct piece of world knowledge like (WORKS-IN MANAGER OFFICE). In section 5.2.2 we will see how this sort of relation might be handled.

2.5. Types of reference

Having reviewed the different sorts of anaphora in English, we are now in a position to make another elaboration of our definition of anaphora. We will distinguish between IDENTITY OF SENSE ANAPHORA[53] (ISA) and IDENTITY OF REFERENCE ANAPHORA (IRA).[54]

An IRA is an anaphor which denotes the same entity as its antecedent. For example, in (2-108)

> (2-108)   Ross made a gherkin sandwich and ate <u>it</u>.

it refers to the very same gherkin sandwich that Ross made. An ISA denotes not the same entity as its antecedent, but one of a similar description. Wasow (1975) offers this example:

> (2-109)   The man who gave his paycheck to his wife was wiser than the man who gave <u>it</u> to his mistress.

Clearly, *it* means the second man's paycheck, not the first man's.

Since the meaning of a text may depend on whether an anaphor is an ISA or an IRA, it is necessary for the complete computer NLU system to be able to tell them apart. This requires the use of semantics and world knowledge. In (2-109), we know *it* is an ISA because, we assume, each man has a paycheck, and an item cannot be given independently to two people at once.

Occasionally below, we will follow Partee (1978) in distinguishing between anaphors which function as bound variables and other anaphors. For example, in (2-110):

---

[53]The term is due to Grinder and Postal (1971), who abbreviate it "I - S = A" ['sic].

[54]An alternative terminology (Nash-Webber 1976): ISA are like DESCRIPTIONAL anaphora, and IRA like DENOTATIONAL anaphora.

(2-110)   No child will admit that <u>he</u> is sleepy.

he is a bound variable anaphor which functions as a place-holder for *child*, much as the bound variable *x* does in the logical form (2-111):

(2-111) $\neg(\exists x:\text{child})$. will-admit-sleepiness x

2.6. Ambiguity in anaphora and default antecedents

Many anaphors, like that of (2-112):

(2-112)   Ross told Daryel <u>he</u> had passed the exam.

are ambiguous — *he* could be either Ross or Daryel. However, some which are theoretically ambiguous are in practice not:

(2-113)   Daryel told Ross <u>he</u>$^{(1)}$ was the ugliest person <u>he</u>$^{(2)}$ knew of.

In this example, each occurrence of *he* could mean either Daryel or Ross, giving a total of four readings for the sentence. Yet most people immediately assume that *he*$^{(1)}$ is Ross and *he*$^{(2)}$ is Daryel without even noticing some or all of the other readings.

This indicates that in many cases of ambiguous anaphors there is a PRE-FERRED or DEFAULT ANTECEDENT, which is taken as the correct one in the absence of contraindicating context or knowledge. The qualification is necessary, as a sentence like (2-113) can be disambiguated by context:

(2-114)   Daryel examined his face disapprovingly in the mirror. When Ross asked him what conclusions he came to, Daryel told Ross <u>he</u> was the ugliest person <u>he</u> knew of.

Both *he*s refer to Daryel here.

More examples to convince the doubtful:

(2-115)   BRISBANE — A terrific right rip from Hector Thompson dropped Ross Eadie at Sandgate on Friday night and won <u>him</u> the Australian welterweight boxing title.[55]

No informant to whom I showed this saw any ambiguity. They were clearly using their knowledge of boxing to infer, without realising it, that it was Thompson (and not Eadie) who won the boxing title. To see that world knowledge is

---

[55]From: *The Canberra times*, 25 May 1977.

the key factor here, we need only consider this report on the sport of dropping, the object of which is to be the first one dropped:

(2-116)   BRISBANE — A terrific right rip from Hector Thompson dropped Ross Eadie at Sandgate on Friday night and won <u>him</u> the Australian welterweight dropping title.

Not all ambiguous anaphors have a default; this one probably doesn't:

(2-117)   SALEM, Ore. — Police Chief Paul Arritola of nearby Jordan Valley runs what could be the most profitable radar speed trap on the continent.
　　　　　Documents filed here in connection with suit against him show that he collected $102,117 in traffic fines last year. Under his contract with this community of 210 people, he gets all the revenue, less the state's share and the cost of running his two-man department. In 1978, <u>that</u> worked out to $70,000.
　　　　　Said Jordan Valley Mayor Ed Krupp: "I'd rather have no comment."[56]

There was no consensus among informants as to whether the police chief ended up with $70,000 or $32,000 because of the ambiguity of *that*. The former case was however slightly preferred (and was probably intended by the writer), since the overall theme of the text is the amount of money that the police chief collected.

That there can, however, be a default referent which is neither the subject nor the theme (see also section 4.1) is shown by this example:

(2-118)   The FBI's role is to ensure our country's freedom and be ever watchful of those who threaten <u>it</u>.[57]

Most informants took *it* to be *our country's freedom* or *our country* (these referents having more or less the same meaning in this context, I assume), rather than *the FBI* or *the FBI's role*, which are also semantically plausible referents, and which are, respectively, the theme and the subject. (Of course, there are those who say that all four candidates have more or less the same meaning in this context.)

Defaults will be discussed further in section 6.5.

An anaphor which can be read as both an IRA and an ISA can make a text ambiguous:

(2-119)   Ross likes his hair short, but Daryel likes <u>it</u> long.

It can be Ross's hair, if an IRA, or Daryel's, if an ISA.

---

[56]From: *The Vancouver express*, 9 March 1979, page A5.

[57]Slightly modified from: Sherman, Craig. [Letter]. *Time*, *111*(20), 15 May 1978.

2.6 Ambiguity in anaphora and default antecedents

Ambiguity may arise only after another anaphor is resolved. The text (2-120) (after Grinder and Postal 1971):

(2-120) Ross loves his wife and Daryel <u>does</u> too.

is ambiguous as to whose wife Daryel loves — his own or Ross's; that is, when *does* is macro-expanded (Hirst 1976b) as *loves his wife*, the *his* is ambiguous.[58] This phenomenon is called SLOPPY IDENTITY.[59]

Sometimes, ambiguities can be resolved by simple lexical information. For example, (2-121) is NOT ambiguous in the same way that (2-120) is, simply because it is inherent in nose twitching that one can only do it to one's own nose:

(2-121) Nadia was able to twitch her nose and Ross was ϕ too.[60]

Similarly, (2-122) is only two ways ambiguous and not four ways as is (2-112), since both anaphors must be coreferential:

(2-122) Ross told Daryel <u>he</u> was able to twitch <u>his</u> nose.

Verb symmetry and reflexivity can also inhibit ambiguity. For example, for all entities A and B *A looks like B* implies *B looks like A*, and *A looks like A* is identically true for all A. Hence (2-123), superficially four ways ambiguous, can only have one meaning, since the two readings with the anaphors coreferential can be dismissed as tautologies (which violate conversational postulates (Gordon and Lakoff 1971; Grice 1975)) and the other two readings are semantically identical:

(2-123) People like dogs because <u>they</u> look like <u>them</u>.

We would not want an NLU system to waste time (or infinitely loop) trying to decide if "people look like dogs" is better than "dogs look like people".

What does all this portend for a computer NLU system? Clearly, it sets certain minimum requirements. A system will need:

1 knowledge about words and their uses;

2 world knowledge;

---

[58]The sentence is unambiguous if we happen to know that Daryel is not married.

[59]Related to the sloppy identity problem is the problem of MISSING ANTECEDENTS, described by Grinder and Postal (1971) who provide this example:

(i) My uncle doesn't have a spouse, but your aunt does and <u>he</u> is lying on the floor.

The referent of *he* is clearly *your aunt's spouse*. This can only be resolved after the ISA proverb *does* is properly interpreted or macro-expanded as *has a spouse*.

[60]Even if Ross had the power to make Nadia's nose twitch, by Pavlovian conditioning for example, we could not express this fact with (2-121), instead having to say something like (i):

(i) Nadia was able to make her nose twitch and Ross was ϕ too

2.6 Ambiguity in anaphora and default antecedents

2 a method of determining default antecedents; and

4 inference mechanisms to apply to 1—3 above and to the meaning of the discourse itself.

2.7. Summary and discussion

In this chapter, I have tried to do these things:

1 define with reasonable precision what anaphora and reference are;

2 give examples of various types of anaphora;

3 demonstrate that a referent can be almost anything in the listener's consciousness, be it explicit or implicit in the discourse, or not in the discourse at all; and

4 show how and why anaphora and reference can be a problem for NLU by computer, and how they are interrelated with other problems in NLU;

5 show that anaphor resolution requires world knowledge, word meaning, inference and default referents.

This chapter, then, has been essentially the statement of a problem. The rest of this thesis looks at solutions to the problem. Because of the fuzziness of the boundary between anaphora and paraphrase and other forms of reference, the problem has, unfortunately, a very fuzzy boundary. It follows by definition that any general resolver of definite reference (clearly a desirable AI goal) will contain an anaphor resolver as a subset. It does not follow, however, that any anaphor resolver can be expanded into a definite reference resolver. Perhaps what is needed is not a happily independent anaphor resolver, but a more general solution to the problem of reference. However, such a solution may not exist, and even if it does, it may not be accessible to us in the near future. Therefore, an independent anaphor resolver is a good step to take next. In subsequent chapters, we shall sometimes, where it is fairly easy to do so, be general and address the problem of reference. At other times, we shall concentrate more particularly on anaphora. This is what AI workers call the vacillation paradigm.