18 DISCOURSE*

Gracie: Oh yeah... And then Mr. and Mrs. Jones were having matrimonial trouble, and my brother was hired to watch Mrs. Jones.
George: Well, I imagine she was a very attractive woman.
Gracie: She was, and my brother watched her day and night for six months.
George: Well, what happened?
Gracie: She finally got a divorce.
George: Mrs. Jones?
Gracie: No, my brother's wife.

George Burns and Gracie Allen in *The Salesgirl*

Up to this point of the book, we have focused primarily on language phenomena that operate at the word or sentence level. Of course, language does not normally consist of isolated, unrelated sentences, but instead of collocated, related groups of sentences. We refer to such a group of sentences as a **discourse**. DISCOURSE

The chapter you are now reading is an example of a discourse. It is in fact a discourse of a particular sort: a **monologue**. Monologues are characterized by a *speaker* (a term which will be used to include writers, as it is here), and a *hearer* (which, analogously, includes readers). The communication flows in only one direction in a monologue, that is, from the speaker to the hearer. MONOLOGUE

After reading this chapter, you may have a conversation with a friend about it, which would consist of a much freer interchange. Such a discourse is called a **dialogue**. In this case, each participant periodically takes turns DIALOGUE

*This chapter by Andrew Kehler

being a speaker and hearer. Unlike a typical monologue, dialogues generally consist of many different types of communicative acts: asking questions, giving answers, making corrections, and so forth.

HCI

Finally, computer systems exist and continue to be developed that allow for *human-computer interaction*, or **HCI**. HCI has properties that distinguish it from normal human-human dialogue, in part due to the present-day limitations on the ability of computer systems to participate in free, unconstrained conversation. A system capable of HCI will often employ a strategy to constrain the conversation in ways that allow it to understand the user's utterances within a limited context of interpretation.

While many discourse processing problems are common to these three forms of discourse, they differ in enough respects that different techniques have often been used to process them. This chapter focuses on techniques commonly applied to the interpretation of monologues; techniques for dialogue interpretation and HCI will be described in Chapter 19.

Language is rife with phenomena that operate at the discourse level. Consider the discourse shown in example (18.1).

(18.1) John went to Bill's car dealership to check out an Acura Integra. He looked at it for about an hour.

What do pronouns such as *he* and *it* denote? No doubt that the reader had little trouble figuring out that *he* denotes John and not Bill, and that *it* denotes the Integra and not Bill's car dealership. On the other hand, toward the end of the exchange presented at the beginning of this chapter, it appears that George had some trouble figuring out who Gracie meant when saying *she*.

What differentiates these two examples? How do hearers interpret discourse (18.1) with such ease? Can we build a computational model of this process? These are the types of questions we address in this chapter. In Section 18.1, we describe methods for interpreting *referring expressions* such as pronouns. We then address the problem of establishing the *coherence* of a discourse in Section 18.2. Finally, in Section 18.3 we explain methods for determining the *structure* of a discourse.

Because discourse-level phenomena are ubiquitous in language, algorithms for resolving them are essential for a wide range of language applications. For instance, interactions with query interfaces and dialogue interpretation systems like ATIS (see Chapter 9) frequently contain pronouns and similar types of expressions. So when a user spoke passage (18.2) to an ATIS system, (18.2) I'd like to get from Boston to San Francisco, on either December 5th or December 6th. It's okay if it stops in another city along the way.

the system had to figure out that *it* denotes the flight that the user wants to book in order to perform the appropriate action.

Similarly, information extraction systems (see Chapter 15) must frequently extract information from utterances that contain pronouns. For instance, if an information extraction system is confronted with passage (18.3),

(18.3) First Union Corp is continuing to wrestle with severe problems unleashed by a botched merger and a troubled business strategy. According to industry insiders at Paine Webber, their president, John R. Georgius, is planning to retire by the end of the year.

it must correctly identify *First Union Corp* as the denotation of *their* (as opposed to *Paine Webber*, for instance) in order to extract the correct event.

Likewise, many text summarization systems employ a procedure for selecting the important sentences from a source document and using them to form a summary. Consider, for example, a news article that contains passage (18.3). Such a system might determine that the second sentence is important enough to be included in the summary, but not the first. However, the second sentence contains a pronoun that is dependent on the first sentence, so it cannot place the second sentence in the summary without first determining the pronoun's denotation, as the pronoun would otherwise likely receive a different interpretation within the summary. Similarly, natural language generation systems (see Chapter 20) must have adequate models for pronominalization to produce coherent and interpretable discourse. In short, just about any conceivable language processing application requires methods for determining the denotations of pronouns and related expressions.

18.1 REFERENCE RESOLUTION

In this section we study the problem of **reference**, the process by which speakers use expressions like *John* and *he* in passage (18.1) to denote a person named John. Our discussion requires that we first define some terminology. A natural language expression used to perform reference is called a **referring expression**, and the entity that is referred to is called the **referent**. Thus, *John* and *he* in passage (18.1) are referring expressions, and John is their referent. (To distinguish between referring expressions and their referents, we italicize the former.) As a convenient shorthand, we will sometimes

REFERENCE

REFERRING EXPRESSION

REFERENT

speak of a referring expression referring to a referent, e.g., we might say that *he* refers to John. However, the reader should keep in mind that what we really mean is that the speaker is performing the act of referring to John by uttering *he*. Two referring expressions that are used to refer to the same entity are said to **corefer**, thus *John* and *he* corefer in passage (18.1). There is also a term for a referring expression that licenses the use of another, in the way that the mention of *John* allows John to be subsequently referred to using *he*. We call *John* the **antecedent** of *he*. Reference to an entity that has been previously introduced into the discourse is called **anaphora**, and the referring expression used is said to be **anaphoric**. In passage (18.1), the pronouns *he* and *it* are therefore anaphoric.

COREFER

ANTECEDENT

ANAPHORA

ANAPHORIC

Natural languages provide speakers with a variety of ways to refer to entities. Say that your friend has an Acura Integra automobile and you want to refer to it. Depending on the operative **discourse context**, you might say *it, this, that, this car, that car, the car*, *the Acura*, *the Integra*, or *my friend's car*, among many other possibilities. However, you are not free to choose between any of these alternatives in any context. For instance, you cannot simply say *it* or *the Acura* if the hearer has no prior knowledge of your friend's car, it has not been mentioned before, and it is not in the immediate surroundings of the discourse participants (i.e., the **situational context** of the discourse).

DISCOURSE CONTEXT

SITUATIONAL CONTEXT

The reason for this is that each type of referring expression encodes different signals about the place that the speaker believes the referent occupies within the hearer's set of beliefs. A subset of these beliefs that has a special status form the hearer's mental model of the ongoing discourse, which we call a **discourse model** (Webber, 1978). The discourse model contains representations of the entities that have been referred to in the discourse and the relationships in which they participate. Thus, there are two components required by a system to successfully produce and interpret referring expressions: a method for constructing a discourse model that evolves with the dynamically-changing discourse it represents, and a method for mapping between the signals that various referring expressions encode and the hearer's set of beliefs, the latter of which includes this discourse model.

DISCOURSE MODEL

We will speak in terms of two fundamental operations to the discourse model. When a referent is first mentioned in a discourse, we say that a representation for it is **evoked** into the model. Upon subsequent mention, this representation is **accessed** from the model. The operations and relationships are illustrated in Figure 18.1.

EVOKED

ACCESSED

We will restrict our discussion to reference to entities, although dis-

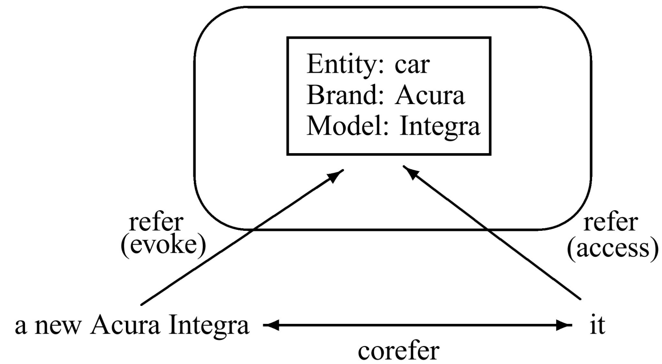

Figure 18.1 Reference operations and relationships. **THIS PICTURE TO BE REPLACED.**

courses include reference to many other types of referents. Consider the possibilities in example (18.4), adapted from Webber (1991).

(18.4) According to John, Bob bought Sue an Integra, and Sue bought Fred a Legend.

a. But *that* turned out to be a lie.
b. But *that* was false.
c. *That* struck me as a funny way to describe the situation.
d. *That* caused Sue to become rather poor.
e. *That* caused them both to become rather poor.

The referent of *that* is a speech act (see Chapter 19) in (18.4a), a proposition in (18.4b), a manner of description in (18.4c), an event in (18.4d), and a combination of several events in (18.4e). The field awaits the development of robust methods for interpreting these types of reference.

Reference Phenomena

The set of referential phenomena that natural languages provide is quite rich indeed. In this section, we provide a brief description of several basic reference phenomena. We first survey five types of referring expression: *indefinite noun phrases*, *definite noun phrases*, *pronouns*, *demonstratives*, and *one-anaphora*. We then describe three types of referents that complicate the reference resolution problem: *inferrables*, *discontinuous sets*, and *generics*.

Indefinite Noun Phrases Indefinite reference introduces entities that are new to the hearer into the discourse context. The most common form of indefinite reference is marked with the determiner *a* (or *an*), as in (18.5), but it can also be marked by a quantifier such as *some* (18.6) or even the determiner *this* (18.7).

(18.5) I saw an Acura Integra today.

(18.6) Some Acura Integras were being unloaded at the local dealership today.

(18.7) I saw this awesome Acura Integra today.

Such noun phrases evoke a representation for a new entity that satisfies the given description into the discourse model.

The indefinite determiner *a* does not indicate whether the entity is identifiable to the speaker, which in some cases leads to a *specific*/*non-specific* ambiguity. Example (18.5) only has the specific reading, since the speaker has a particular Integra in mind, particularly the one she saw. In sentence (18.8), on the other hand, both readings are possible.

(18.8) I am going to the dealership to buy an Acura Integra today.

That is, the speaker may already have the Integra picked out (specific), or may just be planning to pick one out that is to her liking (nonspecific). The readings may be disambiguated by a subsequent referring expression in some contexts; if this expression is definite then the reading is specific (*I hope they still have it*), and if it is indefinite then the reading is nonspecific (*I hope they have a car I like*). This rule has exceptions, however; for instance definite expressions in certain modal contexts (*I will park it in my garage*) are compatible with the nonspecific reading.

Definite Noun Phrases Definite reference is used to refer to an entity that is identifiable to the hearer, either because it has already been mentioned in the discourse context (and thus is represented in the discourse model), it is contained in the hearer's set of beliefs about the world, or the uniqueness of the object is implied by the description itself.

The case in which the referent is identifiable from discourse context is shown in (18.9).

(18.9) I saw an Acura Integra today. *The Integra* was white and needed to be washed.

Examples in which the referent is either identifiable from the hearer's set of beliefs or is inherently unique are shown in (18.10) and (18.11) respectively.

(18.10) *The Indianapolis 500* is the most popular car race in the US.

(18.11) *The fastest car in the Indianapolis 500* was an Integra.

Definite noun phrase reference requires that an entity be accessed from either the discourse model or the hearer's set of beliefs about the world. In the latter case, it also evokes a representation of the referent into the discourse model.

Pronouns Another form of definite reference is pronominalization, illustrated in example (18.12).

(18.12) I saw an Acura Integra today. *It* was white and needed to be washed.

The constraints on using pronominal reference are stronger than for full definite noun phrases, requiring that the referent have a high degree of activation or **salience** in the discourse model. Pronouns usually (but not always) refer to entities that were introduced no further than one or two sentences back in the ongoing discourse, whereas definite noun phrases can often refer further back. This is illustrated by the difference between sentences (18.13d) and (18.13d'). SALIENCE

(18.13) a. John went to Bob's party, and parked next to a beautiful Acura Integra.
b. He went inside and talked to Bob for more than an hour.
c. Bob told him that he recently got engaged.
d. ?? He also said that he bought *it* yesterday.
d.' He also said that he bought *the Acura* yesterday.

By the time the last sentence is reached, the Integra no longer has the degree of salience required to allow for pronominal reference to it.

Pronouns can also participate in **cataphora**, in which they are mentioned before their referents are, as in example (18.14). CATAPHORA

(18.14) Before he bought it, John checked over the Integra very carefully.

Here, the pronouns *he* and *it* both occur *before* their referents are introduced.

Pronouns also appear in quantified contexts in which they are considered to be **bound**, as in example (18.15). BOUND

(18.15) Every woman bought her Acura at the local dealership.

Under the relevant reading, *her* does not refer to some woman in context, but instead behaves like a variable bound to the quantified expression *every woman*. We will not be concerned with the bound interpretation of pronouns in this chapter.

Demonstratives Demonstrative pronouns, like *this* and *that*, behave somewhat differently that simple definite pronouns like *it*. They can appear either alone or as determiners, for instance, *this Acura*, *that Acura*. The choice between two demonstratives is generally associated with some notion of spatial proximity: *this* indicating closeness and *that* signaling distance. Spatial distance might be measured with respect to the discourse participants' situational context, as in (18.16).

(18.16) [John shows Bob an Acura Integra and a Mazda Miata]
Bob (pointing): I like *this* better than *that*.

Alternatively, distance can be metaphorically interpreted in terms of conceptual relations in the discourse model. For instance, consider example (18.17).

(18.17) I bought an Integra yesterday. It's similar to the one I bought five years ago. *That one* was really nice, but I like *this one* even better.

Here, *that one* refers to the Acura bought five years ago (greater temporal distance), whereas *this one* refers to the one bought yesterday (closer temporal distance).

One Anaphora *One*-anaphora, exemplified in (18.18), blends properties of definite and indefinite reference.

(18.18) I saw no less than 6 Acura Integras today. Now I want *one*.

This use of *one* can be roughly paraphrased by *one of them*, in which *them* refers to a plural referent (or generic one, as in the case of (18.18), see below), and *one* selects a member from this set. Thus, *one* may evoke a new entity into the discourse model, but it is necessarily dependent on an existing referent for the description of this new entity.

This use of *one* should be distinguished from the formal, non-specific pronoun usage in (18.19), and its meaning as the number one in (18.20).

(18.19) One shouldn't pay more than twenty thousand dollars for an Acura.

(18.20) John has two Acuras, but I only have one.

Inferrables Now that we have described several types of referring expressions, we now turn our attention to a few interesting types of referents that complicate the reference resolution problem. For instance, in some cases a referring expression does not refer to an entity that has been explicitly evoked in the text, but instead one that is inferentially related to an evoked entity. Such referents are called *inferrables* (Haviland and Clark, 1974; Prince, 1981). Consider the expressions *a door* and *the engine* in sentence (18.21).

(18.21) I almost bought an Acura Integra today, but *a door* had a dent and *the engine* seemed noisy.

The indefinite noun phrase *a door* would normally introduce a new door into the discourse context, but in this case the hearer is to infer something more: that it is not just any door, but one of the doors of the Integra. Similarly, the use of the definite noun phrase *the engine* normally presumes that an engine has been previously evoked or is otherwise uniquely identifiable. Here, no engine has been explicitly mentioned, but the hearer infers that the referent is the engine of the previously mentioned Integra.

Inferrables can also specify the results of processes described by utterances in a discourse. Consider the possible follow-ons (a-c) to sentence (18.22) in the following recipe (from Webber and Baldwin (1992)):

(18.22) Mix the flour, butter, and water.
a. Kneed *the dough* until smooth and shiny.
b. Spread *the paste* over the blueberries.
c. Stir *the batter* until all lumps are gone.

Any of the expressions *the dough* (a solid), *the batter* (a liquid), and *the paste* (somewhere in between) can be used to refer to the result of the actions described in the first sentence, but all imply different properties of this result.

Discontinuous Sets In some cases, references using plural referring expressions like *they* and *them* (see page 672) refer to sets of entities that are evoked together, for instance, using another plural expression (*their Acuras*) or a conjoined noun phrase (*John and Mary*):

(18.23) John and Mary love their Acuras. They drive them all the time.

However, plural references may also refer to sets of entities that have been evoked by discontinuous phrases in the text:

(18.24) John has an Acura, and Mary has a Mazda. They drive them all the time.

Here, *they* refers to John and Mary, and likewise *them* refers to the Acura and the Mazda. Note also that the second sentence in this case will generally receive what is called a *pairwise* or *respectively* reading, in which John drives the Acura and Mary drives the Mazda, as opposed to the reading in which they both drive both cars.

Generics Making the reference problem even more complicated is the existence of *generic* reference. Consider example (18.25).

(18.25) I saw no less than 6 Acura Integras today. *They* are the coolest cars.

Here, the most natural reading is not the one in which *they* refers to the particular 6 Integras mentioned in the first sentence, but instead to the class of Integras in general.

Syntactic and Semantic Constraints on Coreference

Having described a variety of reference phenomena that are found in natural language, we can now consider how one might develop algorithms for identifying the referents of referential expressions. One step that needs to be taken in any successful reference resolution algorithm is to filter the set of possible referents on the basis of certain relatively hard-and-fast constraints. We describe some of these constraints here.

Number Agreement Referring expressions and their referents must agree in number; for English, this means distinguishing between *singular* and *plural* references. A categorization of pronouns with respect to number is shown in Figure 18.2.

Singular	Plural	Unspecified
she, her, he, him, his, it	we, us, they, them	you

Figure 18.2 Number agreement in the English pronominal system.

The following examples illustrate constraints on number agreement.

(18.26) John has a new Acura. It is red.

(18.27) John has three new Acuras. They are red.

(18.28) * John has a new Acura. They are red.

(18.29) * John has three new Acuras. It is red.

Person and Case Agreement English distinguishes between three forms of person: first, second, and third. A categorization of pronoun types with respect to person is shown in Figure 18.3.

The following examples illustrate constraints on person agreement.

(18.30) You and I have Acuras. We love them.

(18.31) John and Mary have Acuras. They love them.

(18.32) * John and Mary have Acuras. We love them. (where *We*=John and Mary)

(18.33) * You and I have Acuras. They love them. (where *They*=You and I)

	First	Second	Third
Nominative	I, we	you	he, she, they
Accusative	me, us	you	him, her, them
Genitive	my, our	your	his, her, their

Figure 18.3 Person and case agreement in the English pronominal system/

In addition, English pronouns are constrained by case agreement; different forms of the pronoun may be required when placed in subject position (nominative case, e.g., *he, she, they*), object position (accusative case, e.g., *him, her, them*), and genitive position (genitive case, e.g., *his Acura, her Acura, their Acura*). This categorization is also shown in Figure 18.3.

Gender Agreement Referents also must agree with the gender specified by the referring expression. English third person pronouns distinguish between *male*, *female*, and *nonpersonal* genders, and unlike many languages, the first two only apply to animate entities. Some examples are shown in Figure 18.4.

masculine	feminine	nonpersonal
he, him, his	she, her	it

Figure 18.4 Gender agreement in the English pronominal system.

The following examples illustrate constraints on gender agreement.

(18.34) John has an Acura. He is attractive. (he=John, not the Acura)

(18.35) John has an Acura. It is attractive. (it=the Acura, not John)

Syntactic Constraints Reference relations may also be constrained by the syntactic relationships between a referential expression and a possible antecedent noun phrase when both occur in the same sentence. For instance, the pronouns in all of the following sentences are subject to the constraints indicated in brackets.

(18.36) John bought himself a new Acura. [himself=John]

(18.37) John bought him a new Acura. [him≠John]

(18.38) John said that Bill bought him a new Acura. [him≠Bill]

(18.39) John said that Bill bought himself a new Acura. [himself=Bill]

(18.40) He said that he bought John a new Acura. [He≠John;he≠John]

English pronouns such as *himself*, *herself*, and *themselves* are called **reflexives**. Oversimplifying the situation considerably, a reflexive corefers REFLEXIVES with the subject of the most immediate clause that contains it (ex. 18.36), whereas a nonreflexive cannot corefer with this subject (ex. 18.37). That this rule applies only for the subject of the most immediate clause is shown by examples (18.38) and (18.39), in which the opposite reference pattern is manifest between the pronoun and the subject of the higher sentence. On the other hand, a full noun phrase like *John* cannot corefer with the subject of the most immediate clause nor with a higher-level subject (ex. 18.40).

Whereas these syntactic constraints apply to a referring expression and a particular potential antecedent noun phrase, these constraints actually prohibit coreference between the two regardless of any other available antecedents that denote the same entity. For instance, normally a nonreflexive pronoun like *him* can corefer with the subject of the previous sentence as it does in example (18.41), but it cannot in example (18.42) because of the existence of the coreferential pronoun *he* in the second clause.

(18.41) John wanted a new car. Bill bought him a new Acura. [him=John]

(18.42) John wanted a new car. He bought him a new Acura. [he=John;him≠John]

The rules given above oversimplify the situation in a number of ways, and there are many cases that they do not cover. Indeed, upon further inspection the facts actually get quite complicated. In fact, it is unlikely that all of the data can be explained using only syntactic relations (Kuno, 1987). For instance, the reflexive *himself* and the nonreflexive *him* in sentences (18.43) and (18.44) respectively can both refer to the subject *John*, even though they occur in identical syntactic configurations.

(18.43) John set the pamphlets about Acuras next to himself. [himself=John]

(18.44) John set the pamphlets about Acuras next to him. [him=John]

For the algorithms discussed later in this chapter, however, we will assume a syntactic account of restrictions on intrasentential coreference.

Selectional Restrictions The selectional restrictions that a verb places on its arguments (see Chapter 16) may be responsible for eliminating referents, as in example (18.45).

(18.45) John parked his Acura in the garage. He had driven it around for hours.

There are two possible referents for *it*, the Acura and the garage. The verb *drive*, however, requires that its direct object denote something that can be driven, such as a car, truck, or bus, but not a garage. Thus, the fact that the pronoun appears as the object of *drive* restricts the set of possible referents to the Acura. It is conceivable that a practical NLP system would include a reasonably comprehensive set of selectional constraints for the verbs in its lexicon.

Selectional restrictions can be violated in the case of metaphor (see Chapter 16); for example, consider example (18.46).

(18.46) John bought a new Acura. It drinks gasoline like you would not believe.

While the verb *drink* does not usually take an inanimate subject, its metaphorical use here allows *it* to refer to *a new Acura*.

Of course, there are more general semantic constraints that may come into play, but these are much more difficult to encode in a comprehensive manner. Consider passage (18.47).

(18.47) John parked his Acura in the garage. It is incredibly messy, with old bike and car parts lying around everywhere.

Here the referent of *it* is almost certainly the garage, but only because a car is probably too small to have bike and car parts laying around 'everywhere'. Resolving this reference requires that a system have knowledge about how large cars typically are, how large garages typically are, and the typical types of objects one might find in each. On the other hand, one's knowledge about Beverly Hills might lead one to assume that the Acura is indeed the referent of *it* in passage (18.48).

(18.48) John parked his Acura in downtown Beverly Hills. It is incredibly messy, with old bike and car parts lying around everywhere.

In the end, just about any knowledge shared by the discourse participants might be necessary to resolve a pronoun reference. However, due in part to the vastness of such knowledge, practical algorithms typically do not rely on it heavily.

Preferences in Pronoun Interpretation

In the previous section, we discussed relatively strict constraints that algorithms should apply when determining possible referents for referring expressions. We now discuss some more readily violated *preferences* that algorithms can be made to account for. These preferences have been posited to apply to pronoun interpretation in particular. Since the majority of work on reference resolution algorithms has focused on pronoun interpretation, we will similarly focus on this problem in the remainder of this section.

Recency Most theories of reference incorporate the notion that entities introduced in recent utterances are more salient than those introduced from utterances further back. Thus, in example (18.49), the pronoun *it* is more likely to refer to the Legend than the Integra.

(18.49) John has an Integra. Bill has a Legend. Mary likes to drive it.

Grammatical Role Many theories specify a salience hierarchy of entities that is ordered by the grammatical position of the expressions which denote them. These invariably treat entities mentioned in subject position as more salient than those in object position, which are in turn more salient than those mentioned in subsequent positions.

Passages such as (18.50) and (18.51) lend support for such a hierarchy. Although the first sentence in each case expresses roughly the same propositional content, the preferred referent for the pronoun *him* varies with the subject in each case – John in (18.50) and Bill in (18.51). In example (18.52), the references to John and Bill are conjoined within the subject position. Since both seemingly have the same degree of salience, it is unclear to which the pronoun refers.

(18.50) John went to the Acura dealership with Bill. He bought an Integra. [he = John]

(18.51) Bill went to the Acura dealership with John. He bought an Integra. [he = Bill]

(18.52) John and Bill went to the Acura dealership. He bought an Integra. [he = ??].

Repeated Mention Some theories incorporate the idea that entities that have been focused on in the prior discourse are more likely to continue to be focused on in subsequent discourse, and hence references to them are more likely to be pronominalized. For instance, whereas the pronoun in example (18.51) has Bill as its preferred interpretation, the pronoun in the final sentence of example (18.53) is more likely to refer to John.

(18.53) John needed a car to get to his new job. He decided that he wanted something sporty. Bill went to the Acura dealership with him. He bought an Integra. [he = John]

Parallelism There are also strong preferences that appear to be induced by parallelism effects, as in example (18.54).

(18.54) Mary went with Sue to the Acura dealership. Sally went with her to the Mazda dealership. [her = Sue]

The grammatical role hierarchy described above ranks Mary as more salient than Sue, and thus should be the preferred referent of *her*. Furthermore, there is no semantic reason that Mary cannot be the referent. Nonetheless, *her* is instead understood to refer to Sue.

This suggests that we might want a heuristic which says that non-subject pronouns prefer non-subject referents. However, such a heuristic may not work for cases that lack the structural parallelism of example (18.54), such as example (18.55), in which Mary is the preferred referent of the pronoun instead of Sue.

(18.55) Mary went with Sue to the Acura dealership. Sally told her not to buy anything. [her = Mary]

Verb Semantics Certain verbs appear to place a semantically-oriented emphasis on one of their argument positions, which can have the effect of biasing the manner in which subsequent pronouns are interpreted. Compare sentences (18.56) and (18.57).

(18.56) John telephoned Bill. He lost the pamphlet on Acuras.

(18.57) John criticized Bill. He lost the pamphlet on Acuras.

These examples differ only in the verb used in the first sentence, yet the subject pronoun in passage (18.56) is typically resolved to John, whereas the pronoun in passage (18.57) is resolved to Bill. Some researchers have claimed that this effect results from what has been called the 'implicit causality' of a verb: the implicit cause of a 'criticizing' event is considered to be its object, whereas the implicit cause of a 'telephoning' event is considered to be its subject. This emphasis results in a higher degree of salience for the entity in this argument position, which leads to the different preferences for examples (18.56) and (18.57).

Similar preferences have been articulated in terms of the thematic roles (see Chapter 16) that the potential antecedents occupy. For example, most hearers resolve *He* to John in example (18.58) and to Bill in example (18.59). Although these referents are evoked from different grammatical role positions, they both fill the Goal thematic role of their corresponding verbs, whereas the other potential referent fills the Source. Likewise, hearers generally resolve *He* to John and Bill in examples (18.60) and (18.61) respectively, providing evidence that fillers of the Stimulus role are preferred over fillers of the Experiencer role.

(18.58) John seized the Acura pamphlet from Bill. He loves reading about cars. (Goal=John, Source=Bill)

(18.59) John passed the Acura pamphlet to Bill. He loves reading about cars. (Goal=Bill, Source=John)

(18.60) The car dealer admired John. He knows Acuras inside and out. (Stimulus=John, Experiencer=the car dealer)

(18.61) The car dealer impressed John. He knows Acuras inside and out. (Stimulus=the car dealer, Experiencer=John)

An Algorithm for Pronoun Resolution

None of the algorithms for pronoun resolution that have been proposed to date successfully account for all of these preferences, let alone succeed in resolving the contradictions that will arise between them. However, Lappin and Leass (1994) describe a straightforward algorithm for pronoun interpretation that takes many of these into consideration. The algorithm employs a simple weighting scheme that integrates the effects of the recency and syntactically-based preferences; no semantic preferences are employed beyond those enforced by agreement. We describe a slightly simplified portion of the algorithm that applies to non-reflexive, third person pronouns.

Broadly speaking, there are two types of operations performed by the algorithm: discourse model update and pronoun resolution. First, when a noun phrase that evokes a new entity is encountered, a representation for it must be added to the discourse model and a degree of salience (which we call a **salience value**) computed for it. The salience value is calculated as the sum of the weights assigned by a set of **salience factors**. The salience factors used and their corresponding weights are shown in Figure 18.5.

SALIENCE VALUE

SALIENCE FACTORS

Sentence recency	100
Subject emphasis	80
Existential emphasis	70
Accusative (direct object) emphasis	50
Indirect object and oblique complement emphasis	40
Non-adverbial emphasis	50
Head noun emphasis	80

Figure 18.5 Salience factors in Lappin and Leass's system.

The weights that each factor assigns to an entity in the discourse model are cut in half each time a new sentence is processed. This, along with the added effect of the sentence recency weight (which initially assigns a weight of 100, to be cut in half with each succeeding sentence), captures the Recency preference described on page 676, since referents mentioned in the current sentence will tend to have higher weights than those in the previous sentence, which will in turn be higher than those in the sentence before that, and so forth.

Similarly, the next five factors in Figure 18.5 can be viewed as a way of encoding a grammatical role preference scheme using the following hierarchy:

> subject $>$ existential predicate nominal $>$ object $>$ indirect object or oblique $>$ demarcated adverbial PP

These five positions are exemplified by the position of the italicized phrases in examples (18.62)–(18.66) respectively.

(18.62) *An Acura Integra* is parked in the lot. (subject)

(18.63) There is *an Acura Integra* parked in the lot. (existential predicate nominal)

(18.64) John parked *an Acura Integra* in the lot. (object)

(18.65) John gave *his Acura Integra* a bath. (indirect object)

(18.66) Inside *his Acura Integra*, John showed Susan his new CD player. (demarcated adverbial PP)

The preference against referents in demarcated adverbial PPs (i.e., those separated by punctuation, as with the comma in example (18.66)) is encoded as a positive weight of 50 for every other position, listed as the non-adverbial emphasis weight in Figure 18.5. This ensures that the weight for any referent is always positive, which is necessary so that the effect of halving the weights is always to reduce them.

The head noun emphasis factor penalizes referents which are embedded in larger noun phrases, again by promoting the weights of referents that are not. Thus, the Acura Integra in each of examples (18.62)–(18.66) will receive 80 points for being denoted by a head noun, whereas the Acura Integra in example (18.67) will not, since it is embedded within the subject noun phrase.

(18.67) The owner's manual for *an Acura Integra* is on John's desk.

Each of these factors contributes to the salience of a referent based on the properties of the noun phrase that denotes it. Of course, it could be that several noun phrases in the preceding discourse refer to the same referent, each being assigned a different level of salience, and thus we need a way in which to combine the contributions of each. To address this, Lappin and Leass associate with each referent an equivalence class that contains all of the noun phrases that have been determined to refer to it. The weight that a salience factor assigns to a referent is the highest of the weights it assigns to the members of its equivalence class. The salience weight for a referent is then calculated by summing these weights for each factor. The scope of a salience factor is a sentence, so, for instance, if a potential referent is mentioned in the current sentence as well as the previous one, the sentence recency weight will be factored in for each. (On the other hand, if the same referent is mentioned more than once in the same sentence, this weight will be counted only once.) Thus, multiple mentions of a referent in the prior discourse can potentially increase its salience, which has the effect of encoding the preference for repeated mentions discussed on page 676.

Once we have updated the discourse model with new potential referents and recalculated the salience values associated with them, we are ready to consider the process of resolving any pronouns that exist within a new sentence. In doing this, we factor in two more salience weights, one for grammatical role parallelism between the pronoun and the potential referent, and one to disprefer cataphoric reference. The weights are shown in Figure 18.6. Unlike the other preferences, these two cannot be calculated independently of the pronoun, and thus cannot be calculated during the discourse model update step. We will use the term *initial salience value* for the weight of a given referent before these factors are applied, and the term *final salience value* for after they have applied.

Role Parallelism	35
Cataphora	-175

Figure 18.6 Per pronoun salience weights in Lappin and Leass's system.

We are now ready to specify the pronoun resolution algorithm. Assuming that the discourse model has been updated to reflect the initial salience values of referents as described above, the steps taken to resolve a pronoun are as follows:

1. Collect the potential referents (up to four sentences back).
2. Remove potential referents that do not agree in number or gender with the pronoun.

3. Remove potential referents that do not pass intrasentential syntactic coreference constraints (as described on page 673).
4. Compute the total salience value of the referent by adding any applicable values from Figure 18.6 to the existing salience value previously computed during the discourse model update step (i.e., the sum of the applicable values in Figure 18.5).
5. Select the referent with the highest salience value. In the case of ties, select the closest referent in terms of string position (computed without bias to direction).

We illustrate the operation of the algorithm by stepping through example (18.68).

(18.68) John saw a beautiful Acura Integra at the dealership. He showed it to Bob. He bought it.

We first process the first sentence to collect potential referents and compute their initial salience values. The following table shows the contribution to salience from each of the salience factors.

	Rec	Subj	Exist	Obj	Ind-Obj	Non-Adv	Head N	Total
John	100	80				50	80	310
Integra	100			50		50	80	280
dealership	100					50	80	230

There are no pronouns to be resolved in this sentence, so we move on to the next, degrading the above values by a factor of two as shown in the following table. The *phrases* column shows the equivalence class of referring expressions for each referent.

Referent	Phrases	Value
John	{ *John* }	155
Integra	{ *a beautiful Acura Integra* }	140
dealership	{ *the dealership* }	115

The first noun phrase in the second sentence is the pronoun *he*. Because *he* specifies male gender, Step 2 of the resolution algorithm reduces the set of possible referents to include only John, so we can stop there and take this to be the referent.

The discourse model must now be updated. First, the pronoun *he* is added in the equivalence class for John. Since *he* occurs in the current sentence and *John* in the previous one, the salience factors do not overlap between the two. The pronoun is in the current sentence (recency=100), subject position (=80), not in an adverbial (=50), and not embedded (=80), and so a total of 310 is added to the current weight for John:

Referent	Phrases	Value
John	{ *John*, he_1 }	465
Integra	{ *a beautiful Acura Integra* }	140
dealership	{ *the dealership* }	115

The next noun phrase in the second sentence is the pronoun *it*, which is compatible with the Integra or the dealership. We first need to compute the final salience values by adding the applicable weights from Figure 18.6 to the initial salience values above. Neither referent assignment would result in cataphora, so that factor does not apply. For the parallelism preference, both *it* and *a beautiful Acura Integra* are in object position within their respective sentences (whereas *the dealership* is not), so a weight of 35 is added to this option. With the Integra having a weight of 175 and the dealership a weight of 115, the Integra is taken to be the referent.

Again, the discourse model must now be updated. Since *it* is in a nonembedded object position, it receives a weight of 100+50+50+80=280, and is added to the current weight for the Integra.

Referent	Phrases	Value
John	{ *John*, he_1 }	465
Integra	{ *a beautiful Acura Integra*, it_1 }	420
dealership	{ *the dealership* }	115

The final noun phrase in the second sentence is *Bob*, which introduces a new discourse referent. Since it occupies an oblique argument position, it receives a weight of 100+40+50+80=270.

Referent	Phrases	Value
John	{ *John*, he_1 }	465
Integra	{ *a beautiful Acura Integra*, it_1 }	420
Bob	{ *Bob* }	270
dealership	{ *the dealership* }	115

Now we are ready to move on to the final sentence. We again degrade the current weights by one half.

Referent	Phrases	Value
John	{ *John*, he_1 }	232.5
Integra	{ *a beautiful Acura Integra*, it_1 }	210
Bob	{ *Bob* }	135
dealership	{ *the dealership* }	57.5

The reader can confirm that the referent of *he* will be resolved to John, and the referent of *it* to the Integra.

The weights used by Lappin and Leass were arrived at by experimentation on a development corpus of computer training manuals. This algorithm, when combined with several filters not described here, achieved 86% accuracy when applied to unseen test data within the same genre. It is possible that these exact weights may not be optimal for other genres (and even more so for other languages), so the reader may want to experiment with these on training data for a new application or language.

In Exercise 18.7, we consider a version of the algorithm that relies only on a noun phrase identifier (see also Kennedy and Boguraev (1996)). In the next paragraphs, we briefly summarize two other approaches to pronoun resolution.

A Tree Search Algorithm Hobbs (1978b) describes an algorithm for pronoun resolution which takes the syntactic representations of the sentences up to and including the current sentence as input, and performs a search for an antecedent noun phrase on these trees. There is no explicit representation of a discourse model or preferences as in the Lappin and Leass algorithm. However, certain of these preferences are approximated by the order in which the search on syntactic trees is performed.

An algorithm that searches parse trees must also specify a grammar, since the assumptions regarding the structure of syntactic trees will affect the results. A fragment for English that the algorithm uses is given in Figure 18.7. The steps of the algorithm are as follows.

1. Begin at the noun phrase (NP) node immediately dominating the pronoun.
2. Go up the tree to the first NP or sentence (S) node encountered. Call this node X, and call the path used to reach it p.
3. Traverse all branches below node X to the left of path p in a left-to-right, breadth-first fashion. Propose as the antecedent any NP node that is encountered which has an NP or S node between it and X.
4. If node X is the highest S node in the sentence, traverse the surface parse trees of previous sentences in the text in order of recency, the

$$
\begin{aligned}
&S \rightarrow NP\ VP\\
&NP \rightarrow \left\{ \begin{array}{l} (Det)\quad Nominal \quad \left(\left\{ \begin{array}{l} PP \\ Rel \end{array} \right\} \right)^* \\ pronoun \end{array} \right\}\\
&Det \rightarrow \left\{ \begin{array}{l} determiner \\ NP\ \text{'s} \end{array} \right\}\\
&PP \rightarrow preposition\ NP\\
&Nominal \rightarrow noun\ (PP)^*\\
&Rel \rightarrow wh\text{-}word\ S\\
&VP \rightarrow verb\ NP\ (PP)^*
\end{aligned}
$$

Figure 18.7 A grammar fragment for the Tree Search algorithm.

most recent first; each tree is traversed in a left-to-right, breadth-first manner, and when an NP node is encountered, it is proposed as antecedent. If X is not the highest S node in the sentence, continue to step 5.

5. From node X, go up the tree to the first NP or S node encountered. Call this new node X, and call the path traversed to reach it p.
6. If X is an NP node and if the path p to X did not pass through the Nominal node that X immediately dominates, propose X as the antecedent.
7. Traverse all branches below node X to the *left* of path p in a left-to-right, breadth-first manner. Propose any NP node encountered as the antecedent.
8. If X is an S node, traverse all branches of node X to the *right* of path p in a left-to-right, breadth-first manner, but do not go below any NP or S node encountered. Propose any NP node encountered as the antecedent.
9. Go to Step 4.

Demonstrating that this algorithm yields the correct coreference assignments for example (18.68) is left as Exercise 18.3.

As stated, the algorithm depends on complete and correct syntactic structures as input. Hobbs evaluated his approach manually (with respect to both parse construction and algorithm implementation) on one hundred examples from each of three different texts, reporting an accuracy of 88.3%. (The accuracy increases to 91.7% if certain selectional restriction constraints are assumed.) Lappin and Leass encoded a version of this algorithm within their system, and reported an accuracy of 82% on their test corpus. Although this is less than the 86% accuracy achieved by their own algorithm, it should be borne in mind that the test data Lappin and Leass used was from the same genre as their development set, but different than the genres that Hobbs used in developing his algorithm.

A Centering Algorithm As we described above, the Hobbs algorithm does not use an explicit representation of a discourse model. The Lappin and Leass algorithm does, but encodes salience as a weighted combination of preferences. Centering theory (Grosz *et al.*, 1995, henceforth GJW), also has an explicit representation of a discourse model, and incorporates an additional claim: that there is a single entity being 'centered' on at any given point in the discourse which is to be distinguished from all other entities that have been evoked.

There are two main representations tracked in the discourse model. In what follows, take U_n and U_{n+1} to be two adjacent utterances. The *backward looking center* of U_n, denoted as $C_b(U_n)$, represents the entity currently being focused on in the discourse after U_n is interpreted. The *forward looking centers* of U_n, denoted as $C_f(U_n)$, form an ordered list containing the entities mentioned in U_n, all of which could serve as the C_b of the following utterance. In fact, $C_b(U_{n+1})$ is by definition the most highly ranked element of $C_f(U_n)$ mentioned in U_{n+1}. (The C_b of the first utterance in a discourse is undefined.) As for how the entities in the $C_f(U_n)$ are ordered, for simplicity's sake we can use the grammatical role hierarchy encoded by (a subset of) the weights in the Lappin and Leass algorithm, repeated below.[1]

> subject > existential predicate nominal > object > indirect object or oblique > demarcated adverbial PP

Unlike the Lappin and Leass algorithm, however, there are no numerical weights attached to the entities on the list, they are simply ordered relative to each other. As a shorthand, we will call the highest-ranked forward-looking center C_p (for 'preferred center').

We describe a centering-based algorithm for pronoun interpretation due to Brennan *et al.* (1987, henceforth BFP). (See also Walker *et al.* (1994); for alternatives, see Kameyama (1986) and Strube and Hahn (1996), inter alia.) In this algorithm, preferred referents of pronouns are computed from relations that hold between the forward and backward looking centers in adjacent sentences. Four intersentential relationships between a pair of utterances U_n and U_{n+1} are defined depending on the relationship between

[1] This is an extended form of the hierarchy used in Brennan *et al.* (1987), described below.

$C_b(U_{n+1})$, $C_b(U_n)$, and $C_p(U_{n+1})$; these are shown in Figure 18.8.

	$C_b(U_{n+1}) = C_b(U_n)$ or undefined $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	Continue	Smooth-Shift
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	Retain	Rough-Shift

Figure 18.8 Transitions in the BFP algorithm.

The following rules are used by the algorithm.

- Rule 1: If any element of $C_f(U_n)$ is realized by a pronoun in utterance U_{n+1}, then $C_b(U_{n+1})$ must be realized as a pronoun also.
- Rule 2: Transition states are ordered. Continue is preferred to Retain is preferred to Smooth-Shift is preferred to Rough-Shift.

Having defined these concepts and rules, the algorithm is defined as follows.

1. Generate possible C_b-C_f combinations for each possible set of reference assignments
2. Filter by constraints, e.g., syntactic coreference constraints, selectional restrictions, centering rules and constraints
3. Rank by transition orderings

The pronominal referents that get assigned are those which yield the most preferred relation in Rule 2, assuming that Rule 1 and other coreference constraints (gender, number, syntactic, selectional restrictions) are not violated.

Let us step through passage (18.68), repeated below as (18.69), to illustrate the algorithm.

(18.69) John saw a beautiful Acura Integra at the dealership. (U_1)
He showed it to Bob. (U_2)
He bought it. (U_3)

Using the grammatical role hierarchy to order the C_f, for sentence U_1 we get:

$C_f(U_1)$: {John, Integra, dealership}
$C_p(U_1)$: John
$C_b(U_1)$: undefined

Sentence U_2 contains two pronouns: *he*, which is compatible with John, and *it*, which is compatible with the Acura or the dealership. John is by definition $C_b(U_2)$, because he is the highest ranked member of $C_f(U_1)$ mentioned in U_2 (again, he is the only possible referent for *he*). We compare the resulting transitions for each possible referent of *it*. If we assume *it* refers to the Acura, the assignments would be:

$C_f(U_2)$: {John, Integra, Bob}
$C_p(U_2)$: John
$C_b(U_2)$: John
Result: Continue $(C_p(U_2)$=$C_b(U_2)$; $C_b(U_1)$ undefined)

If we assume *it* refers to the dealership, the assignments would be:

$C_f(U_2)$: {John, dealership, Bob}
$C_p(U_2)$: John
$C_b(U_2)$: John
Result: Continue $(C_p(U_2)$=$C_b(U_2)$; $C_b(U_1)$ undefined)

Since both possibilities result in a Continue transition, the algorithm does not say which to accept. For the sake of illustration, we will assume that ties are broken in terms of the ordering on the previous C_f list. Thus, we will take *it* to refer to the Integra instead of the dealership, leaving the current discourse model as represented in the first possibility above.

In sentence U_3, *he* is compatible with either John or Bob, whereas *it* is compatible with the Integra. If we assume *he* refers to John, then John is $C_b(U_3)$ and the assignments would be:

$C_f(U_3)$: {John, Acura}
$C_p(U_3)$: John
$C_b(U_3)$: John
Result: Continue $(C_p(U_3)$=$C_b(U_3)$=$C_b(U_2))$

If we assume *he* refers to Bob, then Bob is $C_b(U_3)$ and the assignments would be:

$C_f(U_3)$: {Bob, Acura}
$C_p(U_3)$: Bob
$C_b(U_3)$: Bob
Result: Smooth-Shift $(C_p(U_3)$=$C_b(U_3)$; $C_b(U_3) \neq C_b(U_2))$

Since a Continue is preferred to a Smooth-Shift per Rule 2, John is correctly taken to be the referent.

The main salience factors that the centering algorithm implicitly incorporates include the grammatical role, recency, and repeated mention preferences. Unlike the Lappin and Leass algorithm, however, the manner in which the grammatical role hierarchy affects salience is indirect, since it is the resulting transition type that determines the final reference assignments. In particular, a referent in a low-ranked grammatical role will be preferred to one in a more highly ranked role if the former leads to a more highly ranked transition. Thus, the centering algorithm may (often, but not always, incorrectly) resolve a pronoun to a referent that other algorithms would consider to be of relatively low salience (Lappin and Leass, 1994; Kehler, 1997a). For instance, in example (18.70),

(18.70) Bob opened up a new dealership last week. John took a look at the Acuras in his lot. He ended up buying one.

the centering algorithm will assign Bob as the referent of the subject pronoun *he* in the third sentence – since Bob is $C_b(U_2)$, this assignment results in a Continue relation whereas assigning John results in a Smooth-Shift relation. On the other hand, the Hobbs and Lappin/Leass algorithms will assign John as the referent.

Like the Hobbs algorithm, the centering algorithm was developed on the assumption that correct syntactic structures are available as input. In order to perform an automatic evaluation on naturally occurring data, the centering algorithm would have to be specified in greater detail, both in terms of how all noun phrases in a sentence are ordered with respect to each other on the C_f list (the current approach only includes nonembedded fillers of certain grammatical roles, generating only a partial ordering), as well as how all pronouns in a sentence can be resolved (e.g., recall the indeterminacy in resolving *it* in the second sentence of example (18.68)).

Walker (1989), however, performed a manual evaluation of the centering algorithm on a corpus of 281 examples distributed over texts from three genres, and compared its performance to the Hobbs algorithm. The evaluation assumed adequate syntactic representations, grammatical role labeling, and selectional restriction information as input. Furthermore, in cases in which the centering algorithm did not uniquely specify a referent, only those cases in which the Hobbs algorithm identified the *correct* one were counted as errors. With this proviso, Walker reports an accuracy of 77.6% for centering and 81.8% for Hobbs. See also Tetreault (1999) for a comparison between several centering-based algorithms and the Hobbs algorithm.

18.2 TEXT COHERENCE

Much of the previous section focussed on the nature of anaphoric reference and methods for resolving pronouns in discourse. Anaphoric expressions have often been called **cohesive devices** (Halliday and Hasan, 1976), since the coreference relations they establish serve to 'tie' different parts of a discourse together, thus making it cohesive. While discourses often contain cohesive devices, the existence of such devices alone does not satisfy a stronger requirement that a discourse must meet, that of being *coherent*. In this section, we describe what it means for a text to be coherent, and computational mechanisms for determining coherence. COHESIVE DEVICES

The Phenomenon

Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book. Do you have a discourse? Almost certainly not. The reason is that these utterances, when juxtaposed, will not exhibit **coherence**. Consider, for example, the difference between passages (18.71) and (18.72). COHERENCE

(18.71) John hid Bill's car keys. He was drunk.

(18.72) ?? John hid Bill's car keys. He likes spinach.

While most people find passage (18.71) to be rather unremarkable, they find passage (18.72) to be odd. Why is this so? Like passage (18.71), the sentences that make up passage (18.72) are well formed and readily interpretable. Something instead seems to be wrong with the fact that the sentences are juxtaposed. The hearer might ask, for instance, what hiding someone's car keys has to do with liking spinach. By asking this, the hearer is questioning the coherence of the passage.

Alternatively, the hearer might try to construct an explanation that makes it coherent, for instance, by conjecturing that perhaps someone offered John spinach in exchange for hiding Bill's car keys. In fact, if we consider a context in which we had known this already, the passage now sounds a lot better! Why is this? This conjecture allows the hearer to identify John's liking spinach as the cause of his hiding Bill's car keys, which would explain how the two sentences are connected. The very fact that hearers try to identify such connections is indicative of the need to establish coherence as part of discourse comprehension.

COHERENCE RELATIONS

The possible connections between utterances in a discourse can be specified as a set of **coherence relations**. A few such relations, proposed by Hobbs (1979a), are given below. The terms S_0 and S_1 represent the meanings of the two sentences being related.

Result: Infer that the state or event asserted by S_0 causes or could cause the state or event asserted by S_1.

(18.73) John bought an Acura. His father went ballistic.

Explanation: Infer that the state or event asserted by S_1 causes or could cause the state or event asserted by S_0.

(18.74) John hid Bill's car keys. He was drunk.

Parallel: Infer $p(a_1, a_2, ...)$ from the assertion of S_0 and $p(b_1, b_2, ...)$ from the assertion of S_1, where a_i and b_i are similar, for all i.

(18.75) John bought an Acura. Bill leased a BMW.

Elaboration: Infer the same proposition P from the assertions of S_0 and S_1.

(18.76) John bought an Acura this weekend. He purchased a beautiful new Integra for 20 thousand dollars at Bill's dealership on Saturday afternoon.

Occasion: A change of state can be inferred from the assertion of S_0, whose final state can be inferred from S_1, or a change of state can be inferred from the assertion of S_1, whose initial state can be inferred from S_0.

(18.77) John bought an Acura. He drove to the ballgame.

A mechanism for identifying coherence could support a number of natural language applications, including information extraction and summarization. For example, discourses that are coherent by virtue of the Elaboration relation are often characterized by a summary sentence followed by one or more sentences adding detail to it, as in passage (18.76). Although there are two sentences describing events in this passage, the fact that we infer an Elaboration relation tells us that the same event is being described in each. A mechanism for identifying this fact could tell an information extraction or summarization system to merge the information from the sentences and produce a single event description instead of two.

An Inference Based Resolution Algorithm

Each coherence relation described above is associated with one or more constraints that must be met for it to hold. How can we apply these constraints? To do this, we need a method for performing inference. Perhaps the most familiar type of inference is **deduction**; recall from Section 14.3 that the central rule of deduction is modus ponens: DEDUCTION

$$\frac{\begin{array}{c}\alpha \Rightarrow \beta \\ \alpha\end{array}}{\beta}$$

An example of modus ponens is the following:

$$\frac{\begin{array}{c}\text{All Acuras are fast.} \\ \text{John's car is an Acura.}\end{array}}{\text{John's car is fast.}}$$

Deduction is a form of **sound inference**: if the premises are true, then the conclusion must be true. SOUND INFERENCE

However, much of language understanding is based on inferences that are not sound. While the ability to draw unsound inferences allows for a greater range of inferences to be made, it can also lead to false interpretations and misunderstandings. A method for such inference is logical **abduction** (Pierce, 1955). The central rule of abductive inference is: ABDUCTION

$$\frac{\begin{array}{c}\alpha \Rightarrow \beta \\ \beta\end{array}}{\alpha}$$

Whereas deduction runs an implication relation forward, abduction runs it backward, reasoning from an effect to a potential cause. An example of abduction is the following:

$$\frac{\begin{array}{c}\text{All Acuras are fast.} \\ \text{John's car is fast.}\end{array}}{\text{John's car is an Acura.}}$$

Obviously, this may be an incorrect inference: John's car may be made by another manufacturer yet still be fast.

In general, a given effect β may have many potential causes α_i. We generally will not want to merely reason from a fact to a *possible* explanation of it, we want to identify the *best* explanation of it. To do this, we need a method for comparing the quality of alternative abductive proofs. There are a variety of strategies one could employ for doing this. One possibility is to use a probabilistic model (Charniak and Goldman, 1988; Charniak and Shimony, 1990), although issues arise in choosing the appropriate space over which to calculate these probabilities, and in finding a way to acquire them given the lack of a corpus of events. Another method is to use a purely heuristic strategy (Charniak and McDermott, 1985, Chapter 10) indexCharniak, E., such as preferring the explanation with the smallest number of assumptions, or choosing the explanation that uses the most specific characteristics of the input. While such heuristics may be easy to implement, they generally prove to be too brittle and limiting. Finally, a more general cost-based strategy can be used which combines features (both positive and negative) of the probabilistic and heuristic approaches. The approach to abductive interpretation we illustrate here, due to Hobbs *et al.* (1993), uses such a strategy. To simplify the discussion, however, we will largely ignore the cost component of the system, keeping in mind that one is nonetheless necessary.

Hobbs *et al.* (1993) apply their method to a broad range of problems in language interpretation; here we focus on its use in establishing discourse coherence, in which world and domain knowledge are used to determine the most plausible coherence relation holding between utterances. Let us step through the analysis that leads to establishing the coherence of passage (18.71). First, we need axioms about coherence relations themselves. Axiom (18.78) states that a possible coherence relation is the Explanation relation; other relations would have analogous axioms.

(18.78) $(\forall e_i, e_j) Explanation(e_i, e_j) \supset CoherenceRel(e_i, e_j)$

The variables e_i and e_j represent the events (or states) denoted by the two utterances being related, and the $\supset$ symbol is used to denote the implication relation. In this axiom and those given below, quantifiers always scope over everything to their right. This axiom tells us that, given that we need to establish a coherence relation between two events, one possibility is to abductively assume that the relation is Explanation.

The Explanation relation requires that the second utterance express the cause of the effect that the first sentence expresses. We can state this as axiom (18.79).

(18.79) $(\forall e_i, e_j) cause(e_j, e_i) \supset Explanation(e_i, e_j)$

In addition to axioms about coherence relations, we also need axioms representing general knowledge about the world. The first axiom we use says that if someone is drunk, then others will not want that person to drive, and that the former causes the latter (for convenience, the state of not wanting is denoted by the *diswant* predicate).

(18.80) $(\forall x, y, e_i) drunk(e_i, x) \supset$
$(\exists e_j, e_k) diswant(e_j, y, e_k) \wedge drive(e_k, x) \wedge cause(e_i, e_j)$

Before we move on, a few notes are in order concerning this axiom and the others we will present. First, axiom (18.80) is stated using universal quantifiers to bind several of the variables, which essentially says that in all cases in which someone is drunk, all people do not want that person to drive. Although we might hope that this is generally the case, such a statement is nonetheless too strong. The way in which this is handled in the Hobbs et al. system is by including an additional relation, called an *etc* predicate, in the antecedent of such axioms. An *etc* predicate represents all the other properties that must be true for the axiom to apply, but which are too vague to state explicitly. These predicates therefore cannot be proven, they can only be assumed at a corresponding cost. Because rules with high assumption costs will be dispreferred to ones with low costs, the likelihood that the rule applies can be encoded in terms of this cost. Since we have chosen to simplify our discussion by ignoring costs, we will similarly ignore the use of *etc* predicates.

Second, each predicate has what may look like an 'extra' variable in the first argument position; for instance, the *drive* predicate has two arguments instead of one. This variable is used to reify the relationship denoted by the predicate so that it can be referred to from argument places in other predicates. For instance, reifying the *drive* predicate with the variable e_k allows us to express the idea of not wanting someone to drive by referring to it in the final argument of the *diswant* predicate.

Picking up where we left off, the second world knowledge axiom we use says that if someone does not want someone else to drive, then they do not want this person to have his car keys, since car keys enable someone to drive.

(18.81) $(\forall x, y, e_j, e_k) diswant(e_j, y, e_k) \wedge drive(e_k, x) \supset$
$(\exists z, e_l, e_m) diswant(e_l, y, e_m) \wedge have(e_m, x, z) \wedge carkeys(z, x) \wedge$
$cause(e_j, e_l)$

The third axiom says that if someone doesn't want someone else to have something, he might hide it from him.

(18.82) $(\forall x,y,z,e_i,e_j) diswant(e_l,y,e_m) \wedge have(e_m,x,z) \supset$
$(\exists e_n) hide(e_n,y,x,z) \wedge cause(e_l,e_n)$

The final axiom says simply that causality is transitive, that is, if e_i causes e_j and e_j causes e_k, then e_i causes e_k.

(18.83) $(\forall e_i,e_j,e_k) cause(e_i,e_j) \wedge cause(e_j,e_k) \supset cause(e_i,e_k)$

Finally, we have the content of the utterances themselves, that is, that John hid Bill's car keys (from Bill),

(18.84) $hide(e_1, john, bill, ck) \wedge carkeys(ck, bill)$

and that someone described using the pronoun 'he' was drunk; we will represent the pronoun with the free variable *he*.

(18.85) $drunk(e_2, he)$

We can now see how reasoning with the content of the utterances along with the aforementioned axioms allows the coherence of passage (18.71) to be established under the Explanation relation. The derivation is summarized in Figure 18.9; the sentence interpretations are shown in boxes. We start by assuming there is a coherence relation, and using axiom (18.78) hypothesize that this relation is Explanation,

(18.86) $Explanation(e_1,e_2)$

which, by axiom (18.79), means we hypothesize that

(18.87) $cause(e_2,e_1)$

holds. By axiom (18.83), we can hypothesize that there is an intermediate cause e_3,

(18.88) $cause(e_2,e_3) \wedge cause(e_3,e_1)$

and we can repeat this again by expanding the first conjunct of (18.88) to have an intermediate cause e_4.

(18.89) $cause(e_2,e_4) \wedge cause(e_4,e_3)$

We can take the *hide* predicate from the interpretation of the first sentence in (18.84) and the second *cause* predicate in (18.88), and, using axiom (18.82), hypothesize that John did not want Bill to have his car keys:

(18.90) $diswant(e_3, john, e_5) \wedge have(e_5, bill, ck)$

From this, the *carkeys* predicate from (18.84), and the second *cause* predicate from (18.89), we can use axiom (18.81) to hypothesize that John does not want Bill to drive:

(18.91) $diswant(e_4, john, e_6) \wedge drive(e_6, bill)$

From this, axiom (18.80), and the second *cause* predicate from (18.89), we can hypothesize that Bill was drunk:

(18.92) $drunk(e_2, bill)$

But now we find that we can 'prove' this fact from the interpretation of the second sentence if we simply assume that the free variable *he* is bound to Bill. Thus, the establishment of coherence has gone through, as we have identified a chain of reasoning between the sentence interpretations – one that includes unprovable assumptions about axiom choice and pronoun assignment – that results in $cause(e_2, e_1)$, as required for establishing the Explanation relationship.

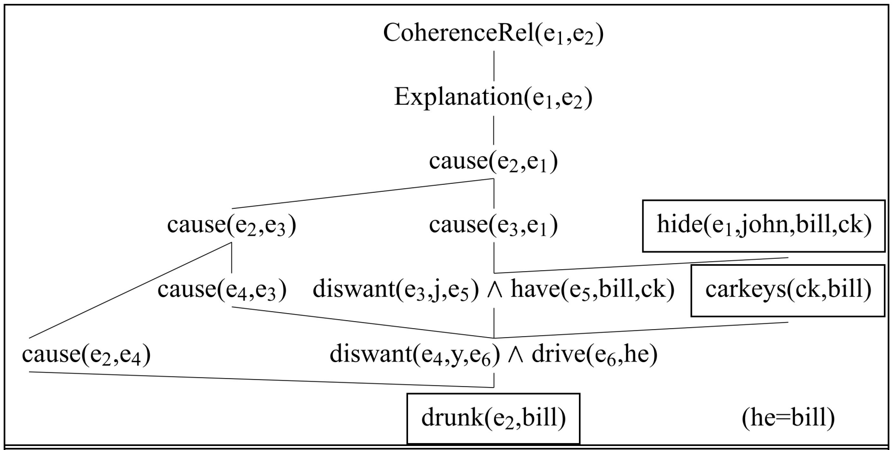

Figure 18.9 Establishing the coherence of passage (18.71).

This derivation illustrates a powerful property of coherence establishment, namely its ability to cause the hearer to infer information about the situation described by the discourse that the speaker has left unsaid. In this case, the derivation required the assumption that John hid Bill's keys because he did not want him to drive (presumably out of fear of him having an accident, or getting stopped by the police), as opposed to some other explanation, such as playing a practical joke on him. This cause is not stated anywhere in passage (18.71); it arises only from the inference process triggered by the need to establish coherence. In this sense, the meaning of a discourse is greater than the sum of the meanings of its parts. That is, a discourse typically communicates far more information than is contained in the interpretations of the individual sentences that comprise it.

We now return to passage (18.72), repeated below as (18.94), which was notable in that it lacks the coherence displayed by passage (18.71), repeated below as (18.93).

(18.93) John hid Bill's car keys. He was drunk.

(18.94) ?? John hid Bill's car keys. He likes spinach.

We can now see why this is: there is no analogous chain of inference capable of linking the two utterance representations, in particular, there is no causal axiom analogous to (18.80) that says that liking spinach might cause someone to not want you to drive. Without additional information that can support such a chain of inference (such as the aforementioned scenario in which someone promised John spinach in exchange for hiding Bill's car keys), the coherence of the passage cannot be established.

Because abduction is a form of unsound inference, it must be possible to subsequently retract the assumptions made during abductive reasoning, that is, abductive inferences are **defeasible**. For instance, if passage (18.93) was followed by sentence (18.95),

DEFEASIBLE

(18.95) Bill's car isn't here anyway; John was just playing a practical joke on him.

the system would have to retract the original chain of inference connecting the two clauses in (18.93), and replace it with one utilizing the fact that the hiding event was part of a practical joke.

In a more general knowledge base designed to support a broad range of inferences, we would probably want axioms that are more general that those we used to establish the coherence of passage (18.93). For instance, consider axiom (18.81), which says that if you do not want someone to drive, then you do not want them to have their car keys. A more general form of the axiom would say that if you do not want someone to perform an action, and an object enables them to perform that action, then you do not want them to have the object. The fact that car keys enable someone to drive would then be encoded separately, along with many other similar facts. Likewise, axiom (18.80) says that if someone is drunk, you don't want them to drive. We might replace this with an axiom that says that if someone does not want something to happen, then they don't want something that will likely cause it to happen. Again, the facts that people typically don't want other people to get into car accidents, and that drunk driving causes accidents, would be encoded separately.

While it is important to have computational models that shed light on the coherence establishment problem, large barriers remain for employing this and similar methods on a wide-coverage basis. In particular, the large number of axioms that would be required to encode all of the necessary facts about the world, and the lack of a robust mechanism for constraining inference with such a large set of axioms, makes these methods largely impractical in practice. Such problems have come to be informally known as **AI-complete**, a play on the term *NP-complete* in computer science. An AI-complete problem is one that essentially requires all of the knowledge – and abilities to utilize it – that humans have. AI-COMPLETE

Other approaches to analyzing the coherence structure of a discourse have also been proposed. One that has received broad usage is Rhetorical Structure Theory (RST) (Mann and Thompson, 1987a), which proposes a set of 23 *rhetorical relations* that can hold between spans of text within a discourse. While RST is oriented more toward text description than interpretation, it has proven to be a useful tool for developing natural language generation systems. RST is described in more detail in Section 20.4.

Coherence and Coreference The reader may have noticed another interesting property of the proof that passage (18.71) is coherent. While the pronoun *he* was initially represented as a free variable, it got bound to Bill during the derivation. In essence, a separate procedure for resolving the pronoun was not necessary; it happened as a side effect of the coherence establishment procedure. In addition to the tree-search algorithm presented on page 683, Hobbs (1978b) proposes this use of the coherence establishment mechanism as a second approach to pronoun interpretation.

This approach provides an explanation for why the pronoun in passage (18.71) is most naturally interpreted as referring to Bill, but the pronoun in passage (18.96) is most naturally interpreted as referring to John.

(18.96) John lost Bill's car keys. He was drunk.

Establishing the coherence of passage (18.96) under Explanation requires an axiom that says that being drunk could cause someone to lose something. Because such an axiom will dictate that the person who is drunk must be the same as the person losing something, the free variable representing the pronoun will become bound to John. The only lexico-syntactic difference between passages (18.96) and (18.71), however, is the verb of the first sentence. The grammatical positions of the pronoun and potential antecedent noun phrases are the same in both cases, so syntactically-based preferences do not distinguish between these.

Discourse Connectives Sometimes a speaker will include a specific cue, called a *connective*, that serves to constrain the set of coherence relations that can hold between two or more utterances. For example, the connective *because* indicates the Explanation relationship explicitly, as in passage (18.97).

(18.97) John hid Bill's car keys because he was drunk.

The meaning of *because* can be represented as $cause(e_2, e_1)$, which would play a similar role in the proof as the *cause* predicate that was introduced abductively via axiom (18.79).

However, connectives do not always constrain the possibilities to a single coherence relation. The meaning of *and*, for instance, is compatible with the Parallel, Occasion, and Result relations introduced on page 690, as exemplified in (18.98)–(18.100) respectively.

(18.98) John bought an Acura and Bill leased a BMW.

(18.99) John bought an Acura and drove to the ballgame.

(18.100) John bought an Acura and his father went ballistic.

However, *and* is not compatible with the Explanation relation; unlike passage (18.97), passage (18.101) cannot mean the same thing as (18.71).

(18.101) John hid Bill's car keys and he was drunk.

While the coherence resolution procedure can use connectives to constrain the range of coherence relations that can be inferred between a pair of utterances, they in and of themselves do not *create* coherence. Any coherence relation indicated by a connective must still be established. Therefore, adding *because* to example (18.72), for instance, still does not make it coherent.

(18.102) ?? John hid Bill's car keys because he likes spinach.

Coherence establishment fails here for the same reason it does for example (18.72), that is, the lack of causal knowledge explaining how liking spinach would cause one to hide someone's car keys.

18.3 Discourse Structure

In the previous section, we saw how the coherence of a pair of sentences can be established. We now ask how coherence can be established for longer discourses. Does one simply establish coherence relations between all adjacent pairs of sentences?

It turns out that the answer is no. Just as sentences have hierarchical structure (that is, syntax), so do discourses. Consider passage (18.103).

(18.103)
- John went to the bank to deposit his paycheck. (S1)
- He then took a train to Bill's car dealership. (S2)
- He needed to buy a car. (S3)
- The company he works for now isn't near any public transportation. (S4)
- He also wanted to talk to Bill about their softball league. (S5)

Intuitively, the structure of passage (18.103) is not linear. The discourse seems to be primarily about the sequence of events described in sentences S1 and S2, whereas sentences S3 and S5 are related most directly to S2, and S4 is related most directly to S3. The coherence relationships between these sentences result in the discourse structure shown in Figure 18.10.

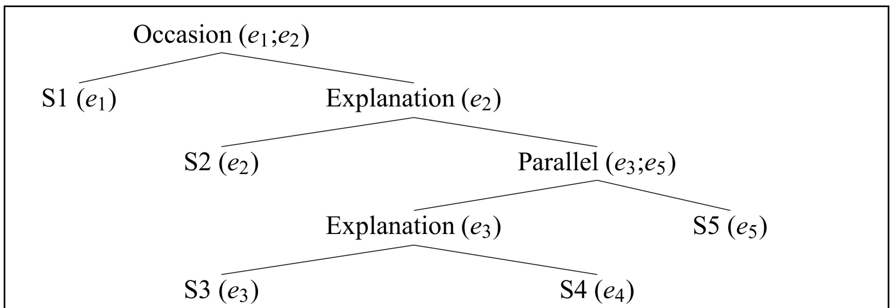

Figure 18.10 The discourse structure of passage (18.103).

Each node in the tree represents a group of locally coherent utterances, called a **discourse segment**. Roughly speaking, one can think of discourse segments as being analogous to intermediate constituents in sentence syntax. DISCOURSE SEGMENT

We can extend the set of discourse interpretation axioms used in the last section to establish the coherence of larger, hierarchical discourses such as (18.103). The recognition of discourse segments, and ultimately discourse structure, results as a by-product of this process.

First, we add axiom (18.104), which states that a sentence is a discourse segment. Here, w is the string of words in the sentence, and e the event or state described by it.

(18.104) $(\forall w, e) sentence(w, e) \supset Segment(w, e)$

Next, we add axiom (18.105), which says that two smaller segments can be composed into a larger one if a coherence relation can be established between the two.

(18.105) $$(\forall w_1, w_2, e_1, e_2, e)\ Segment(w_1, e_1)\ \wedge\ Segment(w_2, e_2) \\ \wedge\ CoherenceRel(e_1, e_2, e) \supset Segment(w_1 w_2, e)$$

Note that extending our axioms for longer discourses has necessitated that we add a third argument to the *CoherenceRel* predicate (e). The value of this variable will be a combination of the information expressed by e_1 and e_2 that represents the main assertion of the resulting segment. For our purposes here, we will assume that **subordinating relations** such as Explanation pass along only one argument (in this case the first, that is, the effect), whereas **coordinating relations** such as Parallel and Occasion pass a combination of both arguments. These arguments are shown in parentheses next to each relation in Figure 18.10.

SUBORDINATING RELATIONS

COORDINATING RELATIONS

Now, to interpret a coherent text W, one must simply prove that it is a segment, as expressed by statement (18.106).

(18.106) $(\exists e) Segment(W, e)$

These two rules will derive any possible binary branching segmental structure for a discourse, as long as that structure can be supported by the establishment of coherence relations between the segments. Herein lies a difference between computing the syntactic structure of a sentence (see Chapter 9) and that of a discourse. Sentence-level grammars are generally complex, encoding many syntactic facts about how different constituents (noun phrases, verb phrases) can modify in each other and in what order. The 'discourse grammar' above, on the contrary, is much simpler, encoding only two rules: a segment rewrites to two smaller segments, and a sentence is a segment. Which of the possible structures is actually assigned depends on how the coherence of the passage is established.

Why would we want to compute discourse structure? Several applications could benefit from it. A summarization system, for instance, might use it to select only the central sentences in the discourse, forgoing the inclusion of subordinate information. For instance, a system for creating brief summaries might only include sentences S1 and S2 when applied to passage (18.103), since the event representations for these were propagated to the top level node. A system for creating more detailed summaries might also include S3 and S5. Similarly, an information retrieval system might weight information in sentences that are propagated to higher-level parts of the discourse structure more heavily than information in ones that are not, and generation systems need knowledge of discourse structure to create coherent discourse, as described in Chapter 20.

Discourse structure may also be useful for natural language subtasks such as pronoun resolution. We already know from Section 18.1 that pronouns display a preference for recency, that is, they have a strong tendency to refer locally. But now we have two possible definitions for recency: recent in terms of the linear order of the discourse, or recent in terms of its hierarchical structure. It has been claimed that the latter definition is in fact the correct one, although admittedly the facts are not completely clear in all cases.

In this section, we have briefly described one of several possible approaches to recovering discourse structure. A different approach, one typically applied to dialogues, will be described in Section 19.4.

18.4 PSYCHOLINGUISTIC STUDIES OF REFERENCE AND COHERENCE

To what extent do the techniques described in this chapter model human discourse comprehension? A substantial body of psycholinguistic research has studied this question.

For instance, a significant amount of work has been concerned with the extent to which people use the preferences described in Section 18.1 to interpret pronouns, the results of which are often contradictory. Clark and Sengal (1979) studied the effects that sentence recency plays in pronoun interpretation using a set of **reading time experiments**. After receiving and acknowledging a three sentence context to read, human subjects were given a target sentence containing a pronoun. The subjects pressed a button when they felt that they understood the target sentence. Clark and Sengal found that the reading time was significantly faster when the referent for the pronoun was evoked from the most recent clause in the context than when it was evoked from two or three clauses back. On the other hand, there was no significant difference between referents evoked from two clauses and three

READING TIME EXPERIMENTS