

An Introduction to Natural Language Generation

Robert Dale
Microsoft Institute of Advanced Software Technology
and
School of Mathematics, Physics, Computing and Electronics
Macquarie University
Sydney
Australia
rdale@microsoft.com

1. An Overview of NLG
2. Linguistic Realization
3. Text Planning
4. Generating Referring Expressions

Topic 4:

Generating Referring Expressions

Overview

- Some terminology and background
- SHRDLU as a simple example
- A framework for generating references
- Pronominalization
- Definiteness and indefiniteness
- Determining the content of a noun phrase
- Generating *one*-anaphora
- Multimodal reference

Generating Referring Expressions

The problem:

the generation of *referring expressions* within connected discourse; and, specifically, the generation of *anaphoric* referring expressions

Generating Referring Expressions

Our principal concerns:

- pronominal anaphora:

A cat walked into the room.

It was wearing a red collar.

- definite noun phrase anaphora:

A cat and a dog walked into the room.

The cat was wearing a red collar.

- *one*-anaphora:

A cat and a dog walked into the room.

The cat was wearing *a red collar*, and the dog was wearing **a blue one**.

Generating Referring Expressions

The issues:

- when can a pronoun be used to refer to an entity?
- when can a definite determiner be used in referring to an entity?
- how is the semantic content of a referring expression decided upon?
- when can *one*-anaphora be used to refer to an entity?

Underlying philosophical questions:

- what does it mean to ‘refer’ to something?
- what kinds of things are ‘entities’?

Definite Reference

- definite noun phrases:

the politician with the Victorian values

- demonstrative terms:

that building over there

- proper names:

Ross Perot

- pronouns:

she

Indefinite Reference

Specific vs. nonspecific indefinite reference:

Every evening **a heron** flies overhead.

1. $\exists y \text{ heron}(y) \wedge [\forall x \text{ evening}(x) \supset \text{flies-overhead}(y)]$
2. $\forall x \text{ evening}(x) \supset [\exists y \text{ heron}(y) \wedge \text{flies-overhead}(y)]$

Reference as a Speech Act

The term ‘referring expression’ is not meant to imply that expressions refer. On the contrary ... reference is a speech act, and speech acts are performed by speakers in uttering words, not by words.

Searle [1969:28]

Reference and Specification

When a speaker uses a referring expression, he or she is specifying an object in his or her discourse model, with the intent of causing the object in question to be introduced or identified in the hearer’s discourse model.

Our Primary Focus

Noun phrases which are:

- anaphoric referring expressions
- specific references
- used to pick out real or hypothetical entities
- in the form of pronouns, definite noun phrases, and *one*-anaphoric noun phrases

Overview

- Some terminology and background
- SHRDLU as a simple example
- A framework for generating references
- Pronominalization
- Definiteness and indefiniteness
- Determining the content of a noun phrase
- Generating *one*-anaphora
- Multimodal reference

Winograd's SHRDLU: An Example Dialogue

User: Stack up both of the red blocks and either a green cube or a pyramid.

SHRDLU: Okay.

User: How did you do it?

SHRDLU: By putting a large red block on the table; then letting go of it; then putting a large green cube on it; then letting go of that cube; then putting the red cube on that cube; then letting go of that cube.

Winograd's SHRDLU: The Knowledge Base

System's knowledge base contains:

- unique objects: I, you, the table, the box, the hand
- other objects by basic classification:
 - (#IS :B1 #BLOCK)
 - (#IS #BLUE #COLOR)

Reference in SHRDLU

To refer to an intended referent x :

- if x is a unique object, then use its name
- otherwise, if x is a colour or shape, use the name of the colour or shape:

What shape is the pyramid?
Pointed.

- otherwise, if x is a block, a ball or a pyramid, then build a noun phrase
- otherwise, fail.

Building NPs in SHRDLU

To refer to x :

- get the noun corresponding to x 's basic classification (**#BLOCK**, **#BALL** or **#PYRAMID**)
- if this does not uniquely describe x , add x 's colour
- if this does not uniquely describe x , add x 's shape
- if the description uniquely describes x , add the determiner *the*
- otherwise, if x supports anything, add *...which supports ...* and the names of the supported objects
- otherwise, add *...which is to the right of ...* and the names of the appropriate objects

Pronominalization in SHRDLU

When to use a pronoun to refer to x :

- if x was referred to as *it* in the question
- if x was referred to as *it* in the question preceding the current one, and the word *it* was not used in the current question
- if x was already named in the current answer, and was the first entity mentioned in the answer
- if no entities have yet been named in the current answer, and x was the only entity named in the previous answer

Pronominalization in SHRDLU

- What color was it?
It was red.
- Did you pick it up?
Yes.
Why?
Because it was on top of the red block.
- How did you do that?
By picking up the yellow block, then putting it on the red block.
- How did you do that?
By picking up the yellow block.
Why?
Because it was in the way.

Discourse Coherency in SHRDLU

- compress lists of similar objects:
yes, four of them: a large blue block, a small red cube, a small red cube, and a small red cube
becomes
yes, four of them: a large blue block and three small red cubes
- *one*-anaphora: compare the current description with the wording of the focus and replace common part with *one*:
Is there a red cube which supports a pyramid?
1. Yes, a large red cube.
2. Yes, a large one.

Discourse Coherency in SHRDLU

- using *that*: if x is already named in the current answer but is not the first-named entity, apply *that* to the appropriate noun:
... by putting a green block on a red cube
then putting that cube in the box

Referring Expressions in SHRDLU

- content determination based on domain heuristics
- pronominalization based on a notion of focus
- surface transformations used to increase coherence

Overview

- Some terminology and background
- SHRDLU as a simple example
- A framework for generating references
- Pronominalization
- Definiteness and indefiniteness
- Determining the content of a noun phrase
- Generating *one*-anaphora
- Multimodal reference

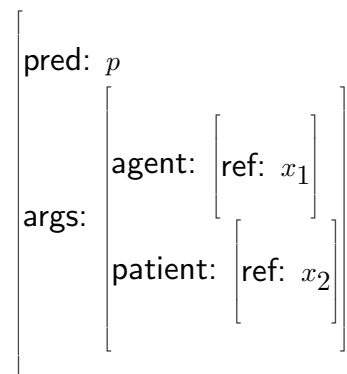
A Framework for Generating References

A discourse consists of a sequence of descriptions of events and states.

- An eventuality specification e is some action, state or event that relates a number of participants x_1, x_2, \dots, x_n
- A discourse specification D is a list of eventuality specifications $\{e_1, e_2, \dots, e_n\}$

A Framework for Generating References

We can assume that each eventuality specification e_i can be represented as a proposition in the following form:



The Knowledge Base

$$\left[\begin{array}{l} \text{index: } t_1 \\ \text{props: } \left[\begin{array}{l} \text{name: table} \end{array} \right] \end{array} \right]$$

$$\left[\begin{array}{l} \text{index: } b_1 \\ \text{props: } \left[\begin{array}{l} \text{shape: cube} \\ \text{colour: red} \\ \text{supports: } b_2 \\ \text{location: ...} \end{array} \right] \end{array} \right]$$

A Framework for Generating References

The discourse model is a triple $\langle C, P, H \rangle$, where:

- C is the set of entities referred to in the current clause (the *current clause contents*)
- P is the set of entities referred to in the current clause (the *previous clause contents*)
- H is a list of all the entities mentioned in the discourse before the previous clause (the *history list*)

Pronominalization and Definiteness

To generate a discourse:

initialize C , P and H to be empty

for each e_i in D

- for each participant x_i in e_i
 - generate a reference to x_i :
 - * if x_i is in C or P then use a pronoun
 - * otherwise if x_i is in H then use a definite noun phrase
 - * otherwise use an indefinite noun phrase
 - add x_i to C
- add contents of P to H ; re-initialize P
- move contents of C to P ; re-initialize C

Overview

- Some terminology and background
- SHRDLU as a simple example
- A framework for generating references
- Pronominalization
- Definiteness and indefiniteness
- Determining the content of a noun phrase
- Generating *one*-anaphora
- Multimodal reference

The Nature of the Referent

- usually an *individual* explicitly referred to previously in the text:

The cat is out.
Who will let him in?

- can be an event (an antecedent NP or clause):

Vina asked Jon to come to *the opening of the exhibition*.

It was going to be a posh affair.

Ewan was buttering the toast while feeding Tom and Elinor with the other hand.

It saves time.

The Semantic Relation of the Referents

- identity
- an implicitly constructed set:

When Joe went to see his solicitor, they spent a rather long time rewriting the contract. When Dom eventually arrived, Joe fooled his solicitor into leaving the room, whereupon he and Dom slipped out the back door. They laughed all the way to the bank.

- explicit mention of the antecedent is not necessary:

Ross sat in the corner, knitting madly. Suddenly he threw it down, and stormed out of the room.

Ross wanted to NAIL the boards together, but Sue made him do it with TAPE.

The Semantic Relation of the Referents

- strained anaphora:

John became a guitarist because he thought that it was a beautiful instrument.

- not all instances of pronouns are anaphoric.

It is raining.

The Location of the Antecedent

- the same sentence or an earlier sentence
- if the antecedent noun phrase is in the same sentence, it may occur either before or after the anaphor:

- in the same clause:

Mary brought *a gallon of orange juice* and drank it in one gulp.

- in a separate clause:

Because she was passing the pet shop, *Judy* was asked to buy a ten kilo bag of cat litter.

- from within a relative clause to the head of that relative clause:

... *a block* which is bigger than anything which supports it

Pronominal Anaphora

Pronominal reference to entities in the preceding discourse:

- Relevant previous research:
 - Sidner [1979]
 - Webber [1979]
 - Grosz, Joshi and Weinstein [1983]
- the LOCALITY CONSTRAINT: if the antecedent is in an earlier sentence, it is usually in the immediately preceding sentence

Tricky Cases

What if there's more than one possible antecedent?

1. • The councillors refused the women a permit because they advocated revolution.
 - The councillors refused the women a permit because they feared revolution.
2. • Sue called Mary a Republican.
 - Then she insulted her.
3. • Sue invited Mary round so that she could cook dinner.

Long Distance Pronominalization

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 rapidly spring 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 her acting that ultimately makes the film.

Sidner's Focus Algorithm

Tracking focus involves:

- the current focus
- the potential focus list: the elements in the current sentence which are candidates for a shift in focus
- the focus stack: elements which have previously been the current focus

Sidner's Focus Algorithm

From the speaker's point of view, there are four options:

- continue talking about the same thing—maintain current focus
- switch to something just introduced—current focus \leftarrow a member of the previous sentence's potential focus list
- return to a previous topic of discussion—current focus \leftarrow popped element from focus stack
- switch to something implicitly related to the current focus—requires general world knowledge to determine that a shift has been made.

Pronominal Anaphora

Uses of Sidner's focus theory in generation:

- McDonald [1980:220]: if the intended referent is the potential actor focus (after Sidner [1979]), this produces a strong vote for a pronoun
- Appelt [1982:129–130]: if x is mutually believed to be in focus, and x is pronominalizable, then use a pronoun; rules are 'adapted from Sidner's rules'
- McKeown [1982:124–132]: focus rules used to decide *what* to say next

Focus in McKeown's TEXT

When faced with a choice as to the next proposition to express: focus of next proposition must be

- same as focus of last proposition; or
- a member of the potential focus list of the last proposition; or
- a member of the focus stack.

So: have to decide whether to

- talk about the same thing; or
- talk about something just introduced; or
- talk about something talked about before.

Focus in McKeown's TEXT

Preference:

- talk about something just introduced rather than what you just talked about; and
- talk about what you just talked about rather than returning to an earlier topic.

Focus in McKeown's TEXT

- each predicate has a default focus argument
- for the ATTRIBUTIVE predicate, this is the first argument:

The chimpanzee has fine control over finger use.
- assumes particular verbs will be used to verbalize the predicate.

Focus in McKeown's TEXT

Focus is used to:

- decide on pronominalization (pronominalize subsequent references to focused object)
- choose between passive and active sentence structures (to keep focused object in the surface subject position)

Overview

- Some terminology and background
- SHRDLU as a simple example
- Pronominalization
- A framework for generating references
- Definiteness and indefiniteness
- Determining the content of a noun phrase
- Generating *one*-anaphora
- Multimodal reference

Definiteness and Indefiniteness

The problem:

deciding whether to use a definite noun phrase or an indefinite noun phrase

Definiteness and Indefiniteness

The naive hypothesis:

entities are introduced into a discourse by means of indefinite reference, whereas subsequent references to already introduced entities are usually of the form of definite referring expressions.

A man and a woman walked into the room.
The man was wearing a funny hat.

However ...

A bus turned the corner.
The driver had a mean look in her eye.

Definiteness and Indefiniteness

Why we should worry:

- the determiners *the* and *a* are amongst the most common words in the English language
- in a corpus analysed by Kučera and Francis [1967], *the* is the most frequently used of 50406 distinct words: it occurs 69971 times in a corpus of 1,014,232 words—i.e., *the* accounts for 6.9% of all words in the corpus
- the word *a* occurs 23237 times, ranking fifth.

The Logical Analysis

Definiteness as existential uniqueness

So

The king of France is bald.

is equivalent to the assertion of the conjunction of the following three propositions:

There is a king of France.
 There is not more than one king of France.
 This individual is bald.

Contextually-defined Domains

A better solution:

- instead of interpreting *the F* as denoting x if and only if x is the one and only F in existence, we should interpret *the F* as denoting x if and only if x is the one and only F in some contextually-defined domain of discourse.

Problem: how do we determine what the required domain of discourse is?

Hawkin's Location Theory of Reference

Shared sets:

PREVIOUS DISCOURSE SET: those entities which have already been mentioned in the current discourse, and any other discourses the current speaker and hearer have participated in

IMMEDIATE SITUATION SET: those entities which are evoked 'situationally'

LARGER SITUATION SET: those entities which are shared by virtue of being part of specific or general background knowledge

ASSOCIATION SET: those entities whose existence is inferrable on the basis of association with entities which are already known

Definiteness and Indefiniteness

Hawkin's view of what a definite referring expression does:

- introduces a referent (or referents) to the hearer
- instructs the hearer to locate the referent in some shared set of objects
- refers to the totality of the objects or mass within this set which satisfy the referring expression.

The use of shared sets solves the problem of the 'scope of uniqueness' introduced by the naive logical analysis.

Salience Rankings

the F denotes *x* if and only if *x* is the most salient *F* in the domain of discourse, according to some contextually determined salience ranking

[Lewis 1979:241]

But what's the relevant notion of salience?
Two general approaches:

- each entity in a discourse can be assigned a value on a continuous scale of **ACTIVATEDNESS**; the entity with the highest activation level is then the most salient
- a two level distinction between **LOCAL FOCUS** and **GLOBAL FOCUS**, analogous to the psychological distinction between short term and long term memory

Grosz and Sidner

Local and global focus:

- a discourse consists of hierarchically arranged **DISCOURSE SEGMENTS**
- each discourse segment may itself consist of a number of discourse segments, and so on
- a series of utterances constitute a discourse segment if they together serve to realize a particular **DISCOURSE PURPOSE**
- corresponding to each discourse segment, there is a **FOCUS SPACE**
- clue words and phrases such as *however*, *in any case*, *finally*, and so on, are generally agreed to be surface indicators of discourse structure

Grosz and Sidner

The structures of discourse structure:

- linguistic structure
- intentional structure
- attentional state

Grosz and Sidner

The interaction of reference and discourse structure:

Just as linguistic devices affect structure, so the discourse segmentation affects the interpretation of linguistic expressions in a discourse. *Referring expressions provide the primary example of this effect.* The segmentation of discourse constrains the use of referring expressions by delineating certain points at which there is a significant change in what entities (objects, properties, or relations) are being discussed. For example, *there are different constraints on the use of pronouns and reduced definite noun phrases within a segment than across segment boundaries.* While discourse segmentation is obviously not the only factor governing the use of referring expressions, it is an important one.

[Grosz and Sidner 1986:178]

Grosz and Sidner

The Tent Example

- | | | |
|-----|-----|---|
| DS0 | 1. | A: I'm going camping next weekend. Do you have a two-person tent I could borrow? |
| | 2. | B: Sure. I have a two-person backpacking tent . |
| DS1 | 3. | A: The last trip I was on there was a huge storm. |
| | 4. | It poured for two hours. |
| | 5. | I had a tent, but I got soaked anyway. |
| | 6. | B: What kind of a tent was it? |
| | 7. | A: A tube tent. |
| | 8. | B: Tube tents don't stand up well in a real storm. |
| | 9. | A: True. |
| | 10. | B: Where are you going on this trip? |
| | 11. | A: Up in the Minarets. |
| | 12. | B: Do you need any other equipment? |
| | 13. | A: No. |
| | 14. | B: Okay. I'll bring the tent in tomorrow. |

Grosz and Sidner

The Movies Example

- | | |
|----|--|
| 1. | The "movies" are so attractive to the great American public, |
| 2. | especially to young people, |
| 3. | that it is time to take careful thought about their effect on mind and morale. |
| 4. | Ought any parent to permit his children to attend a moving picture show often or without being quite certain of the show he permits them to see? ... |

Grosz and Sidner

The Screw Example

18. A: The two screws are loose, but I'm having trouble getting the wheel off.
 19. E: Use [the wheelpuller]. Do you know how to use it?
 20. A: No.
 21. E: Do you know what it looks like?
 22. A: Yes.
 23. E: Show it to me please.
 24. A: OK.
 25. E: Good. Loosen [the screw in the center] and place the jaws around the hub of the wheel, then tighten the screw onto the center of the shaft. . . .

Grosz and Sidner

The Compressor Example

- E: Good morning. I would like for you to re-assemble [the compressor].
 . . .
 E: I suggest you begin by attaching the pump to the platform.
 [. . . other subtasks]
 E: Good. All that remains then is to attach the belt housing cover to the belt housing frame.
 A: All right. I assume the hole in the housing cover opens to the pump pulley rather than to the motor pulley.
 E: Yes, that is correct. The pump pulley also acts as a fan to cool the pump.
 A: All right, the belt housing cover is on and tightened down.
 [30 minutes and 60 utterances after beginning]
 E: Fine. Now let's see if [it] works.

Long Distance Pronominalization

Some observations:

- full definite noun phrases are sometimes used in situations where the antecedent noun phrase is in the previous sentence
- two possible reasons for rarity of long distance pronominalization:
 - all instances of long distance pronominalization are merely sloppy and inconsiderate language use
 - long distance pronominalization is legitimate, but it may only occur under special circumstances

Long Distance Pronominalization

- if the hearer is presented with a pronoun to resolve, this indicates that the discourse segment containing the previous utterance has *not* been closed; so, using a definite noun phrase is a way of indicating that a segment *has* closed
- the only place a pronoun *cannot* be used to refer to an entity in the preceding clause is in the first utterance immediately following the closure of a discourse segment
- the only place it is legitimate to use long distance pronominalization is in this location, since it can't be mistaken for immediate pronominalization

Overview

- Some terminology and background
- SHRDLU as a simple example
- A framework for generating references
- Pronominalization
- Definiteness and indefiniteness
- Determining the content of a noun phrase
- Generating *one*-anaphora
- Multimodal reference

Determining the Content of an NP

The problem:

how do we decide on the semantic content of a noun phrase referring expression?

The principles of reference:

- the principle of sensitivity
- the principle of efficiency
- the principle of adequacy

To meet these principles, we need:

- a model of the hearer
- a model of the discourse

Modelling the Hearer

Knowledge about language:

- the speaker must know what language the hearer understands
- the speaker must know what words and expressions in that language the hearer understands

Knowledge about the world:

- the speaker must know which entities are known to the hearer, besides those which have been explicitly mentioned in the discourse
- the speaker must know what entities the hearer can infer the existence of

Determining the Content of an NP

Discriminatory power:

We view the generation of subsequent referring expressions as distinguishing an entity from that set of entities with which it might be confused. We call this set of potential distractors the **CONTEXT SET**.

Two questions:

- how do we go about distinguishing an entity from a set of other entities?
- how do we determine the constituency of the context set?

Distinguishing Descriptions: The Problem

Given:

- a set of knowledge base entities C
- an intended referent $r \in C$
- a knowledge base of properties P

we require a set of properties in P which are together true of r but of no other entity in C .

- A DISTINGUISHING DESCRIPTION of r is the linguistic realisation of this set of properties.
- A MINIMAL DISTINGUISHING DESCRIPTION of r is the linguistic realisation of the smallest such set of properties.

Computing Distinguishing Descriptions

Three steps which are repeated until a successful description has been constructed:

1. Check whether the description constructed so far is successful in picking out the intended referent from the context set.
2. If it's not sufficient, choose the most useful fact that will contribute to the description.
3. Extend the description with this fact, and reduce the context set accordingly.

Computing Distinguishing Descriptions

Initial Conditions:

- $C_r = \{\langle \text{all entities} \rangle\}$;
- $P_r = \{\langle \text{all properties true of } r \rangle\}$;
- $L_r = \{\}$

1. Check Success

- **if** $|C_r| = 1$ **then** return L_r as a DD
- **elseif** $P_r = 0$ **then** return L_r as a non-DD
- **else goto** Step 2.

2. Choose Property

- **for each** $p_i \in P_r$ **do**:
- $C_{r_i} \leftarrow C_r \cap \{x | p_i(x)\}$
- Chosen property is p_j , where C_{r_j} is smallest set.
- **goto** Step 3.

3. Extend Description (wrt the chosen p_j)

- $L_r \leftarrow L_r \cup \{p_j\}$
- $C_r \leftarrow C_{r_j}$
- $P_r \leftarrow P_r - \{p_j\}$
- **goto** Step 1.

An Example

Suppose we know:

$\{\text{cup}(c_1), \text{cup}(c_2), \text{cup}(c_3),$
 $\text{blue}(c_1), \text{red}(c_2), \text{red}(c_3),$
 $\text{big}(c_1), \text{small}(c_2), \text{big}(c_3)\}$

and we want to refer to c_3 .

Chosen properties:

1. $\text{red}(c_3)$ (or $\text{big}(c_3)$)
2. $\text{big}(c_3)$ (or $\text{red}(c_3)$)

Some Problems

1. The algorithm does not guarantee to find a *minimal* distinguishing description [Reiter 1990].
2. The mechanism doesn't necessarily produce a *useful* description.
3. The algorithm doesn't represent what people seem to do when constructing a referring expression.

Discriminatory Power

A problem:

what if there is more than one minimal distinguishing description?

Possible factors determining choice:

- the purpose of the description
- there may be a conventional order in which properties are selected
- properties may be selected according to some ordering which is context-dependent (for example, global factors such as the particular domain of discourse, and local factors such as properties previously used in the current discourse)

Discriminatory Power

Another problem:

how do we determine the context set?

A simple approach:

- the entire contents of the discourse model—i.e., every entity that has been mentioned in the discourse—is the relevant context

But:

- the larger the context with respect to which an entity must be distinguished, the more computationally expensive the process of distinguishing the intended referent
- need some way of cutting down the size of the context

Discourse Structure

The interaction of reference and discourse structure:

Just as linguistic devices affect structure, so the discourse segmentation affects the interpretation of linguistic expressions in a discourse. *Referring expressions provide the primary example of this effect.* The segmentation of discourse constrains the use of referring expressions by delineating certain points at which there is a significant change in what entities (objects, properties, or relations) are being discussed. For example, *there are different constraints on the use of pronouns and reduced definite noun phrases within a segment than across segment boundaries.* While discourse segmentation is obviously not the only factor governing the use of referring expressions, it is an important one.

[Grosz and Sidner 1986:178]

Discourse Structure

The basic idea:

the entities last mentioned in focus spaces near the top of the focus stack are more salient than those last mentioned in focus spaces further down the focus stack

Discourse Structure

A little more formally ...

- take FS_n to be the set of entities introduced in focus space n , where n is the ordinal position of the focus space on the stack, with the 0th space being the space on the bottom of the stack
- take s to be a function that determines the salience of a set of entities
- then, if FS_T is the focus space on the top of the focus stack, the following holds:

$$s(FS_T) > s(FS_{T-1}) > \dots > s(FS_0)$$

What Do People Do?

Two salient points that emerge from psychological research:

- Human speakers in many cases include unnecessary modifiers in the referring expressions they construct.
- Human speakers can begin to utter a referring expression before they have finished scanning the set of distractors.

Since human speakers include redundant information in their referring expressions, it's unnecessary for a natural language generation system to ensure that its output never contains redundant information.

Some Other Questions

- Faced with a choice, which attribute should be used?
- Is it preferable to add a modifier or to use a more specific head noun?
- Should relative or absolute adjectives be used?

Some Plausible Hypotheses

- Speakers prefer to use adjectives that describe easily perceptible properties such as size, shape, or colour.
- Hearers sometimes have trouble determining if an object belongs to a specialized class, so adding an explicit modifier is better than using a specialized head noun.
- Human speakers seem to prefer to use relative adjectives.

An Incremental Algorithm

Iterate through a task-dependent list of attributes:

- add an attribute to the description being constructed if it rules out any distractors that have not already been ruled out;
- terminate when a distinguishing description has been constructed.

Driving the Incremental Algorithm

- `PreferredAttributes` lists the attributes that human speakers and hearers prefer.
- These attributes should be listed in order of preference, with the most preferred attribute first.
- The elements of this list and their order will vary with the domain, and will typically be determined by empirical investigation.

Assumptions about the Knowledge Base

- Every entity is characterised in terms of a collection of attributes and their values: e.g. `<colour, red>`.
- Every entity has as one of its attributes some type.
- The knowledge base may organize some attribute values in a subsumption taxonomy.

Interface Functions:

- `MoreSpecificValue(object, attribute, value)`
- `BasicLevelValue(object, attribute)`
- `UserKnows(object, attribute-value-pair)`

Inputs and Outputs

The host system must provide:

- a symbol corresponding to the intended referent; and
- a list of symbols corresponding to the members of the contrast set.

The algorithm returns a list of attribute–value pairs that correspond to the semantic content of the referring expression to be realized.

The Algorithm

General approach: the algorithm iterates through the attributes in PreferredAttributes.

- For each attribute, it checks if specifying a value for that attribute would rule out at least one member of the contrast set that has not already been ruled out; if so, this attribute is added.
- Continue until a referring expression has been formed that rules out every member of the contrast set.
- There is no backtracking; once an attribute–value pair has been added to the referring expression, it is not removed even if the addition of subsequent attribute–value pairs make it unnecessary.
- A head noun is always included, even if it has no discriminatory power.

Script-Based Reference

Initial Conditions: $C_r = \langle \text{all entities} \rangle$; $P = \langle \text{preferred attributes} \rangle$; $L_r = \{ \}$

1. Check Success

- if $|C_r| = 1$ then return L_r as a distinguishing description
- elseif $P = 0$ then return L_r as a non-dd
- else goto Step 2.

2. Evaluate Next Property

- get next $p_i \in P$ such that $\text{userknows}(p_i(r))$
- if $|\{x \in C_r | p_i(x)\}| < |C_r|$ then goto Step 3
- else goto Step 2.

3. Extend Description (wrt the chosen p_j)

$L_r \leftarrow L_r \cup \{p_j\}$; $C_r \leftarrow C_{r_j}$; goto Step 1.

Performance

- It is fast. The algorithm's run-time is linear in the number of distractors and independent of the number of potential attributes.
- It allows human preferences and capabilities to be taken into consideration via the Preferred Attributes list and the UserKnows function.
- In use in the IDAS system.

Overview

- Some terminology and background
- SHRDLU as a simple example
- A framework for generating references
- Pronominalization
- Definiteness and indefiniteness
- Determining the content of a noun phrase
- Generating *one*-anaphora
- Multimodal reference

One-Anaphora

- used to identify an individual which is of the same *type* as some other individual mentioned in the discourse
- also referred to as DESCRIPTIONAL anaphora

Different uses of *one*

- substitute *one*:
 Judy bought *a large green plastic wombat*.
 Mary bought one too.
 Judy bought *a small blue plastic wombat*.
 Mary bought a large green one.
- the indefinite personal pronoun
One should never put one's cat out at night.
- the cardinal number *one*:
 I have one cat, although I'd like to have more.

One as Substitute

- the head noun which is replaced must be a count noun:
These biscuits are stale.
 Get some fresh ones.
This bread is stale.
 Get some fresh.
- syntactic structure of the *one*-anaphoric NP and antecedent need not match:
 Do you have *any bullets made of platinum?*
 No, but I have some leaden ones.
- *one*-anaphora is transparent to number:
 Jon bought *a grey T-shirt*.
 Marc preferred the brighter ones.
 Jon bought *some coloured T-shirts*.
 Marc preferred the green one.

One as Substitute

There are *semantic* constraints on possible combinations:

Do you have *a milk bottle*?
 ?No, but I have a red one.

One-Anaphora

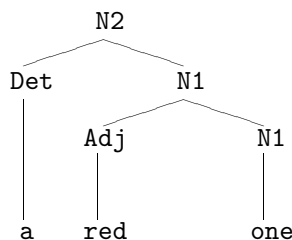
Two kinds of substitution:

Does anyone have *any green wombat covers*?
 Yes, I have some.
 Do you want *the green wombat covers*?
 No, I want the blue ones.

One-Anaphora

The *one/ones* pair function as N1 pro-forms; must always appear with a determiner.

Has anyone got *a blue jumper*?
 No, but I've got a red one.



One-Anaphora

The *one/some* pair can be analysed as pro-N2 forms:

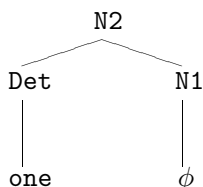
Who has *a copy of Cosmopolitan*?
 I have one.



One-Anaphora

Alternatively, *one/some* can be analysed as instances of the indefinite article:

Glyn brought Mike a debugging aid.
Jo brought Henk two.



An Algorithm for One-Anaphora

Assuming that *one* literally substitutes for N1 syntactic constituents:

- construct the complete syntactic structure for the required semantic content
- compare this structure with that of NPs in the preceding utterance: if they have any N1 constituents in common, find the largest and replace this by the pro-N1 form *one*

The recursive structure of N1 constituents means that this deals with examples like:

Helen bought a large old Germanic manuscript.
Marc could only afford a small one.

One-Anaphora as syntactic substitution

Problems with this simple approach:

- not clear how the rule can be elegantly integrated with a rule for the use of *one/some*
- does not rule out the following:

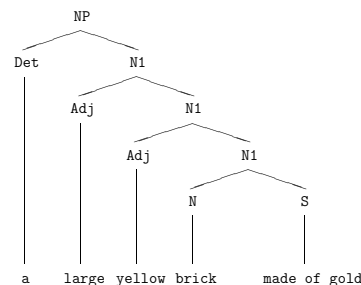
Do you have any wine bottles?
No, but I have a red one.
- this suggests that the substitution should be semantically constrained

One-Anaphora as syntactic substitution

A more serious problem:

Do you have a large yellow brick made of gold?
No, but I have a small *one*.

A possible syntactic structure:



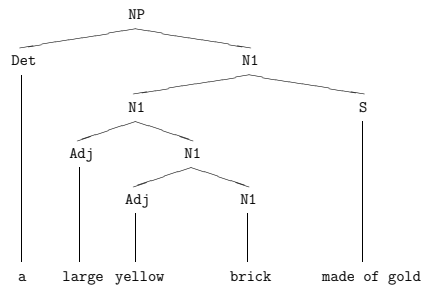
One-Anaphora as syntactic substitution

...but also consider

Do you have a large yellow brick made of gold?

No, but I have a small *one* made of silver.

A possible analysis:



One-Anaphora as syntactic substitution

- no single structural analysis of the original noun phrase makes available all the possible antecedents for *one*-anaphors as syntactic constituents
- so, we need a structure in which all the properties of the entity other than that which is to be realized by the head noun are maintained by means of what is essentially a flat list
- given previous considerations, this should be a *semantic* structure

Overview

- Some terminology and background
- SHRDLU as a simple example
- A framework for generating references
- Pronominalization
- Definiteness and indefiniteness
- Determining the content of a noun phrase
- Generating *one*-anaphora
- Multimodal reference

Multimodality

- we use pointing in conjunction with speech: multimodality in dialogue
- we use pictures or graphics in conjunction with text: multimodality in text

MULTIMODAL DEIXIS: the combination of deictic expressions like *this* and *there* with extralinguistic devices such as pointing gestures.

The XTRA System

- NL access system for expert systems
- developed at Saarbrücken
- domain: assists user in filling out a tax form visible on screen
- elements of form can be specified via either typed verbal descriptions or by combinations of descriptions and simulated pointing gestures

The Advantages of Pointing

- can use shorter, simpler or even referentially insufficient descriptions
- can use pointing when don't know how to describe intended referent

Example:

THIS ↑ is broken

while pointing at some part of the engine of one's car.

Different Kinds of Pointing

PUNCTUAL POINTING: • indicates single point on form

- specifies primitive objects (eg individual regions or entities)
- also reference to complex region by pointing to part

NON-PUNCTUAL POINTING: performance of complex motion, eg underlining something or drawing a border around something

MULTIPLE POINTING: more than one pointing gesture per utterance: eg specifying elements of sets, several examples of one concept

Pointing actions in XTRA

Can refer to:

- a **FORM REGION**: You can enter your donations HERE ↑.
- an **ENTRY**: THESE 350 DM ↑ are travel expenses.
- a correlated **CONCEPT**: Can I deduct SUCH DONATIONS ↑?

Pointing actions in XTRA

- XTRA's generator = Reithinger's POPEL
- POPEL-WHAT is concerned with the givenness, the pointability of an object, and the situation-dependency of description
- POPEL-HOW is concerned with linguistic constraints and the generation of descriptions

Heuristics for Mode Choice

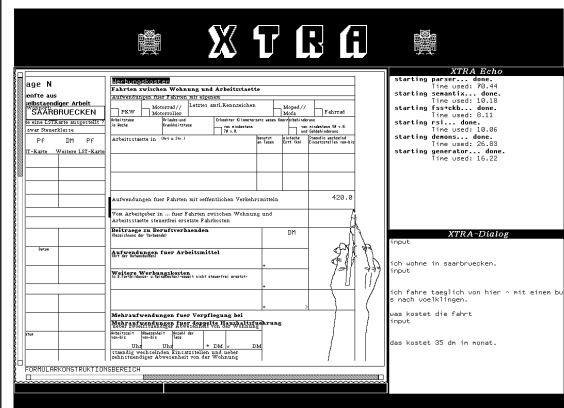
- Sentence generation is performed incrementally
- When a decision as to whether or not to point has to be made, POPEL-WHAT doesn't know the whole content of the sentence
- Heuristics are used:

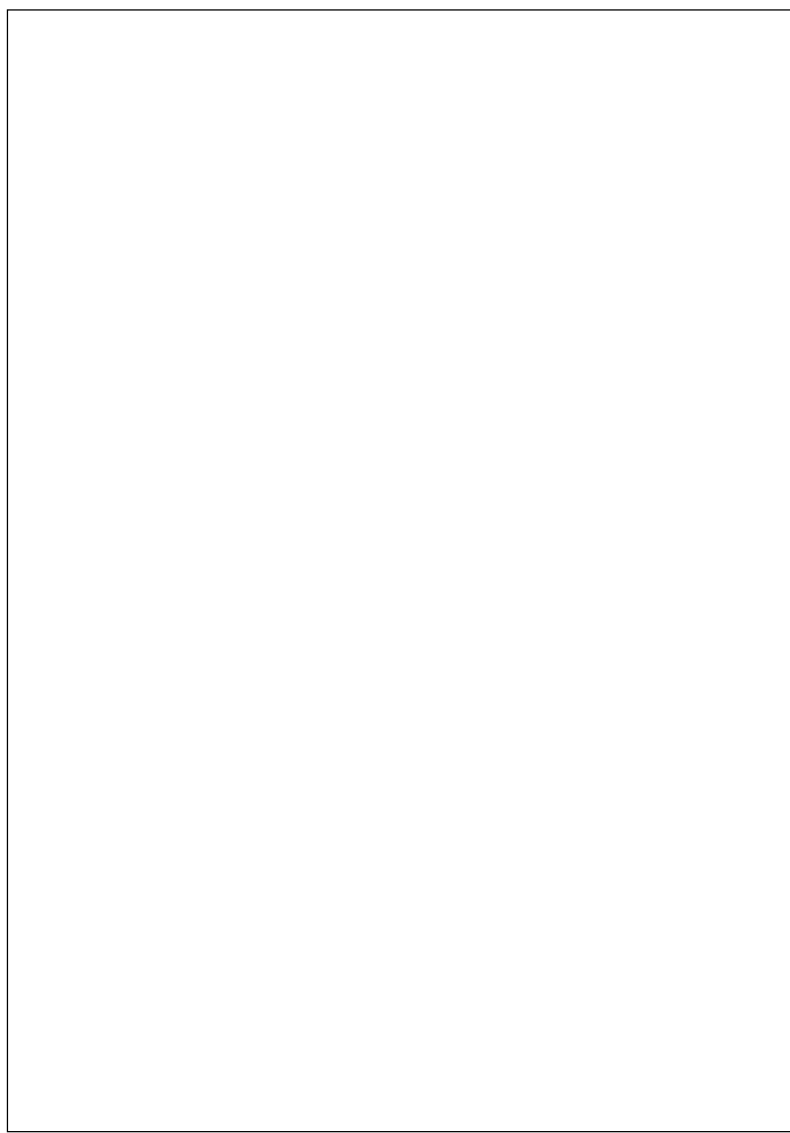
Do not use a pointing gesture if the object in question can also be specified by a short referential expression, eg a pro-word.

Punctual Pointing Gestures

- Disambiguation: a gesture can refer to either a field or its contents so linguistic information is required: *this field* vs *this amount of money*.
- If the pointing action refers to a field, the pencil is in the middle of the field.
- If it refers to an entry, the pencil is below the entry (so the entry isn't obscured).

Generating Pointing Gestures



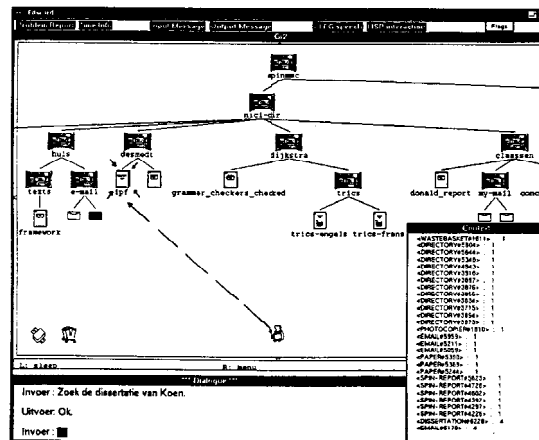


Special Features of Pointing at a Form

Domain limitations which simplify the task:

- structure of the space pointed at: forms are two-dimensional, consist of non-overlapping regions
- type of pointing: TACTILE POINTING—the object the gesture refers to is always touched with the pointing device, so only three possible referents:
 - the region of the form,
 - the entry in a value region,
 - the concept the region is an instance of.

Claassen's EDWARD



Claassen's EDWARD

- Central notion: salience.
- A context factor (CF) is defined by:
 - scope (a collection of individual entities)
 - significance weight (numerical value)
 - decay function (how the significance weight is to be decreased after creation)
- salience value of a concept = the sum of the significance weights of the CFs that have that concept in their scope
- provides a unified measure of salience that brings together independent factors

Linguistic Context Factors in EDWARD

Four Types:

- main term referent CF, initial SW = 3
- subject referent CF, initial SW = 2
- nested term referent CF, initial SW = 1
- relation CF, initial SW = 3

Decay function: subtract 1 at each successive update; if CF = 0, discard CF.

Linguistic CFS: An Example

The author of this article lives in Nijmegen.

Four CFS:

- main term referent CF, scope = {the author of this article, Nijmegen}
- subject referent CF, scope = {the author of this article}
- nested term referent CF, scope = {this article}
- relation CF, scope = {live in, write}

So salience of *the author of this article* is 5.

Perceptual Context Factors

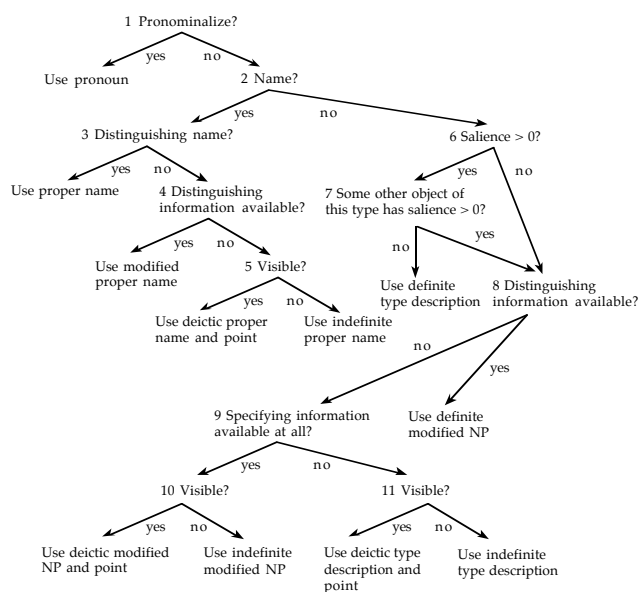
Three types:

- visible referent CF: initial SW = 1, drops to 0 if the icon becomes invisible
- selected referent CF: created when an icon is selected, initial SW = 2, remains 2 as long as selected; drops to 0 when deselected
- indicated referent CF: initial SW = 30; after first update = 1, after second = 0

EDWARD's Conceptual Generator

- determines the type and conceptual content of the linguistic expressions that will be used to refer to the entity and relation instances in the input message
- determines which referents should be pointed at

Choosing a Form of Reference



Generating Referring Expressions

What we've covered:

- pronominalization
- definiteness and indefiniteness
- determining the content of a noun phrase
- *one*-anaphora
- generating multimodal references

Linguistic Realisation

- most widely used approaches based on unification and systemics
- publicly available implementations of both kinds of systems
- all these approaches are more related than might at first seem to be the case

Text Planning

- two main approaches based on grammar-like text schemas and more atomic rhetorical relationships
- ultimately the only difference may be the degree of precompilation
- still no substantive theory of discourse to parallel work in syntactic structure
- scope for extension to multimodal documents

Generating Referring Expressions

- the pronominalisation problem may be AI-hard
- wide range of usable subsequent reference strategies available
- the initial reference problem awaits a better theory of purpose in reference

Some Provocative Quotes

If investigation of communication is seen as the underlying task of research, then generation gives much better access to that task, since it is much easier to develop methods and programs that work with whole discourses rather than being restricted to single sentences or a small number of sentences at best. [Mann 1989]

If comprehension researchers stopped dwelling on parsing and grammar formalisms, they would realize that the bulk of the work remains to be done, and that they are not that much better off for the vastly greater number of person-years that have been put in on the problem. [McDonald 1988]