

COOKING UP REFERRING EXPRESSIONS

Robert Dale

Centre for Cognitive Science, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, Scotland
email: rda%uk.ac.ed.epistemi@nss.cs.ucl.ac.uk

ABSTRACT

This paper describes the referring expression generation mechanisms used in EPICURE, a computer program which produces natural language descriptions of cookery recipes. Major features of the system include: an underlying ontology which permits the representation of non-singular entities; a notion of *discriminatory power*, to determine what properties should be used in a description; and a PATR-like unification grammar to produce surface linguistic strings.

INTRODUCTION

EPICURE (Dale 1989a, 1989b) is a natural language generation system whose principal concern is the generation of referring expressions which pick out complex entities in connected discourse. In particular, the system generates natural language descriptions of cookery recipes. Given a top level goal, the program first decomposes that goal recursively to produce a plan consisting of operations at a level of detail commensurate with the assumed knowledge of the hearer. In order to describe the resulting plan, EPICURE then models its execution, so that the processes which produce referring expressions always have access to a representation of the ingredients in the state they are in at the time of description.

This paper describes that part of the system responsible for the generation of subsequent referring expressions, i.e., references to entities which have already been mentioned in the discourse. The most notable features of the approach taken here are as follows: (a) the use of a sophisticated underlying ontology, to permit the representation of non-singular entities; (b) the use of two levels of semantic representation, in conjunction with a model of the discourse, to produce appropriate anaphoric referring expressions; (c) the use of a notion of *discriminatory power*, to determine what properties should be used in describing an entity; and (d) the use of a PATR-like unification grammar (see, for example, Karttunen (1986); Shieber (1986)) to pro-

duce surface linguistic strings from input semantic structures.

THE REPRESENTATION OF INGREDIENTS

In most natural language systems, it is assumed that all the entities in the domain of discourse are singular individuals. In more complex domains, such as recipes, this simplification is of limited value, since a large proportion of the objects we find are masses or sets, such as those described by the noun phrases *two ounces of salt* and *three pounds of carrots* respectively.

In order to permit the representation of entities such as these, EPICURE makes use of a notion of a generalized physical object or *physobj*. This permits a consistent representation of entities irrespective of whether they are viewed as individuals, masses or sets, by representing each as a *knowledge base entity* (KBE) with an appropriate *structure* attribute. The knowledge base entity corresponding to *three pounds of carrots*, for example, is that shown in figure 1.

A knowledge base entity models a *physobj* in a particular *state*. An entity may change during the course of a recipe, as processes are applied to it: in particular, apart from gaining new properties such as being peeled, chopped, etc., an ingredient's *structure* may change, for example, from set to mass. Each such change of state results in the creation of a new knowledge base entity. Suppose, for example, a grating event is applied to our three pounds of carrots between states s_0 and s_1 : the entity shown in figure 1 will then become a mass of grated carrot, represented in state s_1 by the KBE shown in figure 2.

BUILDING A REFERRING EXPRESSION

To construct a referring expression corresponding to a knowledge base entity, we first build a *deep se-*

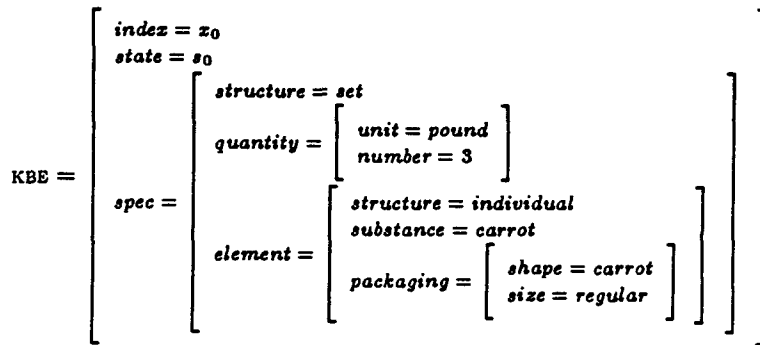


Figure 1: The knowledge base entity corresponding to *three pounds of carrots*

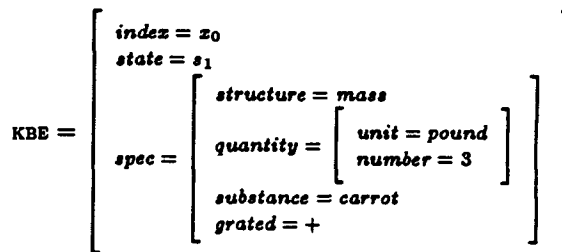


Figure 2: The knowledge base entity corresponding to *three pounds of grated carrot*

semantic structure which specifies the semantic content of the noun phrase to be generated. We call this the *recoverable* semantic content, since it consists of just that information the hearer should be able to derive from the corresponding utterance, even if that information is not stated explicitly: in particular, elided elements and instances of *one-anaphora* are represented in the deep semantic structure by their more semantically complete counterparts, as we will see below.

From the deep semantic structure, a *surface semantic structure* is then constructed. Unlike the deep semantic structure, this closely matches the syntactic structure of the resulting noun phrase, and is suitable for passing directly to a PATR-like unification grammar. It is at the level of surface semantic structure that processes such as elision and *one-anaphora* take place.

PRONOMINALIZATION

When an entity is to be referred to, we first check to see if pronominalization is possible. Some previous approaches to the *pronominalization deci-*

sion have taken into account a large number of contextual factors (see, for example, McDonald (1980:218-220)). The approach taken here is relatively simple. EPICURE makes use of a discourse model which distinguishes two principal components, corresponding to Grosz's (1977) distinction between local focus and global focus. We call that part of the discourse model corresponding to the local focus *cache memory*: this contains the lexical, syntactic and semantic detail of the current utterance being generated, and the same detail for the previous utterance. Corresponding to global focus, the discourse model consists of a number of hierarchically-arranged *focus spaces*, mirroring the structure of the recipe being described. These focus spaces record the semantic content, but not the syntactic or lexical detail, of the remainder of the preceding discourse. In addition, we make use of a notion of *discourse centre*: this is intuitively similar to the notion of centering suggested by Grosz, Joshi and Weinstein (1983), and corresponds to the focus of attention in the discourse. In recipes, we take the centre to be the *result* of

the previous operation described. Thus, after an utterance like *Soak the butterbeans* the centre is the entity described by the noun phrase *the butterbeans*. Subsequent references to the centre can be pronominalized, so that the next instruction in the recipe might then be *Drain and rinse them*.

Following Grosz, Joshi and Weinstein (1983), references to other entities present in cache memory may also be pronominalized, provided the centre is pronominalized.¹

If the intended referent is the current centre, then this is marked as part of the *status* information in the deep semantic structure being constructed, and a null value is specified for the structure's descriptive content. In addition, the verb case frame used to construct the utterance specifies whether or not the linguistic realization of the entity filling each case role is obligatory: as we will see below, this allows us to model a common linguistic phenomenon in recipes (*recipe context empty objects*, after Massam and Roberge (1989)). For a case role whose surface realization is obligatory, the resulting deep semantic structure is then as follows:

$$DS = \left[\begin{array}{l} \text{index} = x \\ \text{sem} = \left[\begin{array}{l} \text{status} = \left[\begin{array}{l} \text{given} = + \\ \text{centre} = + \\ \text{oblig} = + \end{array} \right] \\ \text{spec} = \left[\begin{array}{l} \text{type} = \phi \end{array} \right] \end{array} \right] \end{array} \right]$$

This will be realized as either a pronoun or an elided NP, generated from a surface semantic structure which is constructed in accordance with the following rules:

- If the *status* includes the features [*centre*, +] and [*oblig*, +], then there should be a corresponding element in the surface semantic structure, with a null value specified for the descriptive content of the noun phrase to be generated;

¹We do not permit pronominal reference to entities last mentioned before the previous utterance: support for this restriction comes from a study by Hobbs, who, in a sample of one hundred consecutive examples of pronouns from each of three very different texts, found that 98% of antecedents were either in the same or previous sentence (Hobbs 1978:322-323). However, see Dale (1988) for a suggestion as to how the few instances of *long-distance pronominalization* that do exist might be explained by means of a theory of discourse structure like that suggested by Gross and Sidner (1986).

- If the *status* includes the features [*centre*, +] and [*oblig*, -], then this participant should be omitted from the surface semantic structure altogether.

In the former case, this will result in a pronominal reference as in *Remove them*, where the surface semantic structure corresponding to the pronominal form is as follows:

$$SS = \left[\begin{array}{l} \text{index} = x \\ \text{sem} = \left[\begin{array}{l} \text{status} = \left[\begin{array}{l} \text{given} = + \\ \text{centre} = + \\ \text{oblig} = + \end{array} \right] \\ \text{spec} = \left[\begin{array}{l} \text{agr} = \left[\begin{array}{l} \text{num} = \text{pl} \\ \text{case} = \text{acc} \end{array} \right] \\ \text{desc} = \phi \end{array} \right] \end{array} \right] \end{array} \right]$$

However, if the participant is marked as non-obligatory, then reference to the entity is omitted, as in the following:

Fry the onions.
Add the garlic ϕ .

Here, the case frame for *add* specifies that the indirect object is non-obligatory; since the entity which fills this case role is also the centre, the complete prepositional phrase *to the onions* can be elided. Note, however, that the entity corresponding to *the onions* still figures in the deep semantic structure; thus, it is integrated into the discourse model, and is deemed to be part of the semantic content recoverable by the hearer.

FULL DEFINITE NOUN PHRASE REFERENCE

If pronominalization is ruled out, we have to build an appropriate description of the intended referent. In EPICURE, the process of constructing a description is driven by two principles, very like Gricean conversational maxims (Grice 1975). The *principle of adequacy* requires that a referring expression should identify the intended referent unambiguously, and provide sufficient information to serve the purpose of the reference; and the *principle of efficiency*, pulling in the opposite direction, requires that the referring expression used must not contain more information than is necessary for the task at hand.²

These principles are implemented in EPICURE

²Similar considerations are discussed by Appelt (1985).

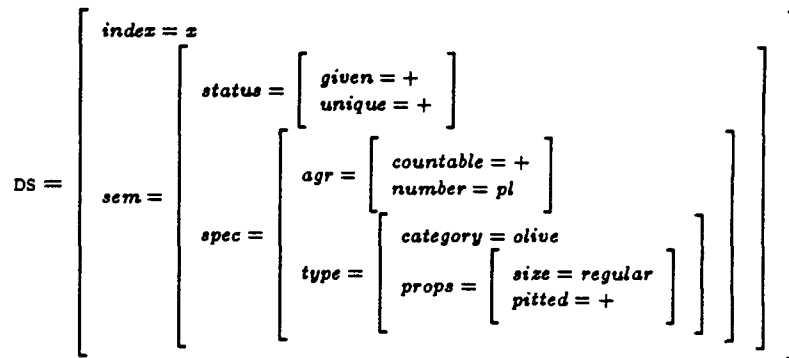


Figure 3: The deep semantic structure corresponding to *the pitted olives*

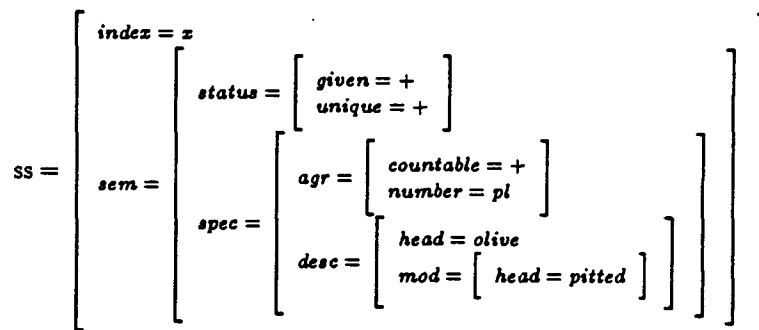


Figure 4: The surface semantic structure corresponding to *the pitted olives*

by means of a notion of *discriminatory power*. Suppose that we have a set of entities U such that

$$U = \{x_1, x_2, \dots, x_n\}$$

and that we wish to distinguish one of these entities, x_i , from all the others. Suppose, also, that the domain includes a number of attributes (a_1, a_2 , and so on), and that each attribute has a number of permissible values (v_{n_1}, v_{n_2} , and so on); and that each entity is described by a set of attribute-value pairs. In order to distinguish x_i from the other entities in U , we need to find some set of attribute-value pairs which are together true of x_i , but of no other entity in U . This set of attribute-value pairs constitutes a *distinguishing description* of x_i with respect to the context U . A *minimal distinguishing description* is then a set of such attribute-value pairs, where the cardinality of that set is such that there are no other sets of attribute-value pairs of lesser cardinality which are sufficient

to distinguish the intended referent.

We find a minimal distinguishing description by observing that different attribute-value pairs differ in the effectiveness with which they distinguish an entity from a set of entities. Suppose U has N elements, where $N > 1$. Then, any attribute-value pair true of the intended referent x_i will be true of n entities in this set, where $n \geq 1$. For any attribute-value pair $\langle a, v \rangle$ that is true of the intended referent, we can compute the discriminatory power (notated here as F) of that attribute-value pair with respect to U as follows:

$$F(\langle a, v \rangle, U) = \frac{N-n}{N-1} \quad 1 \leq n \leq N$$

F thus has as its range the interval $[0,1]$, where a value of 1 for a given attribute-value pair indicates that the attribute-value pair singles out the intended referent from the context, and a value of

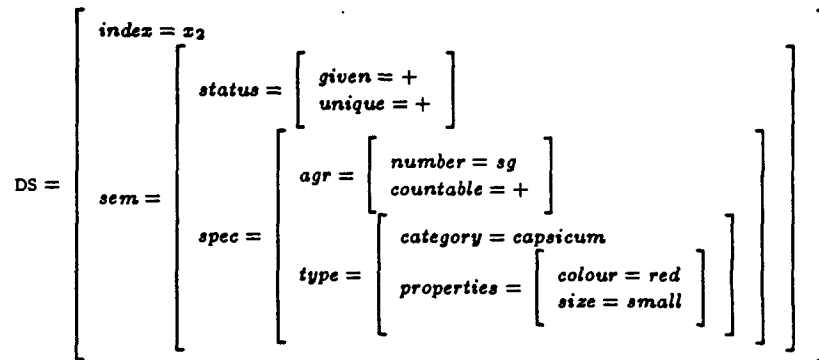


Figure 5: The deep semantic structure corresponding to *the small red capsicum*

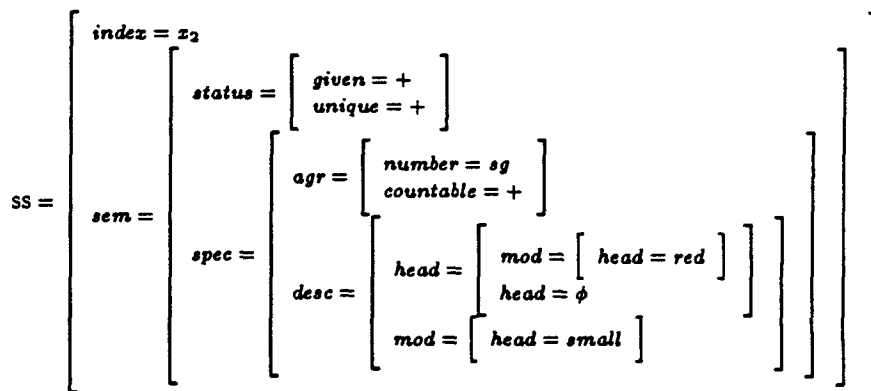


Figure 6: The surface semantic structure corresponding to *the small red one*

0 indicates that the attribute-value pair is of no assistance in singling out the intended referent.

Given an intended referent and a set of entities from which the intended referent must be distinguished, this notion is used to determine which set of properties should be used in building a description which is both adequate and efficient.³ There remains the question of how the constituency of the set U of entities is determined: in the present work, we take the context always to consist of the *working set*. This is the set of distinguishable enti-

³Strictly speaking, this mechanism is only applicable in the form described here to those properties of an entity which are realizable by what are known as *absolute* (or *intersective* or *predicative*) adjectives (see, for example, Kamp (1975), Keenan and Faltz (1978)). This is acceptable in the current domain, where many of the adjectives used are derived from the verbs used to describe processes applied to entities.

ties in the domain at any given point in time: the constituency of this set changes as a recipe proceeds, since entities may be created or destroyed.⁴

Suppose, for example, we determine that we must identify a given object as being a set of olives which have been pitted (in a context, for example, where there are also olives which have not been pitted); the corresponding deep semantic structure is then as in figure 3.

Note that this deep semantic structure can be realized in at least two ways: as either *the olives which have been pitted* or *the pitted olives*.

⁴A slightly more sophisticated approach would be to restrict U to exclude those entities which are, in Grosz and Sidner's (1986) terms, only present in closed focus spaces. However, the benefit gained from doing this (if indeed it is a valid thing to do) is minimal in the current context because of the small number of entities we are dealing with.

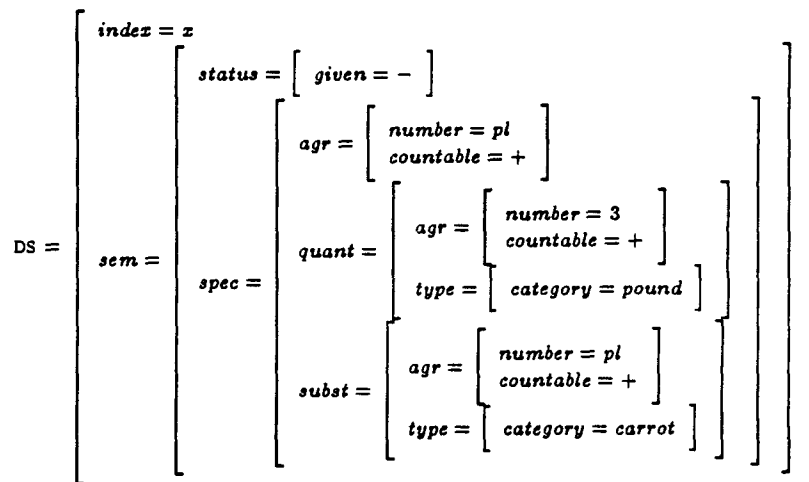


Figure 7: The deep semantic structure corresponding to *three pounds of carrots*

Both forms are possible, although they correspond to different *surface* semantic structures. Thus, the generation algorithm is non-deterministic in this respect (although one might imagine there are other factors which determine which of the two realizations is preferable in a given context). The surface semantic structure for the simpler of the two noun phrase structures is as shown in figure 4.

ONE ANAPHORA

The algorithms employed in EPICURE also permit the generation of *one*-anaphora, as in

Slice the large green capsicum.
Now remove the top of the small red one.

The deep semantic structure corresponding to the noun phrase *the small red one* is as shown in figure 5.

The mechanisms which construct the surface semantic structure determine whether *one*-anaphora is possible by comparing the deep semantic structure corresponding to the previous utterance with that corresponding to the current utterance, to identify any elements they have in common. The two distinct levels of semantic representation play an important role here: in the deep semantic structure, only the basic semantic category of the description has special status (this is similar to Weber's (1979) use of restricted quantification), whereas the embedding of the surface semantic structure's

desc feature closely matches that of the noun phrase to be generated. For *one*-anaphora to be possible, the two deep semantic structures being compared must have the same value for the feature addressed by the path $\langle \text{sem spec type category} \rangle$. Rules which specify the relative ordering of adjectives in the surface form are then used to build an appropriately nested surface semantic structure which, when unified with the grammar, will result in the required *one*-anaphoric noun phrase. In the present example, this results in the surface semantic structure in figure 6.

PSEUDO-PARTITIVE NPS

Partitive and pseudo-partitive noun phrases, exemplified by *half of the carrots* and *three pounds of carrots* respectively, are very common in recipes; EPICURE is capable of generating both. So, for example, the pseudo-partitive noun phrase *three pounds of carrots* (as represented by the knowledge base entity shown in figure 1) is generated from the deep semantic structure shown in figure 7 via the surface semantic structure shown in figure 8.

The generation of partitive noun phrases requires slightly different semantic structures, described in greater detail in Dale (1989b).

THE UNIFICATION GRAMMAR

Once the required surface semantic structure has been constructed, this is passed to a unification

NP	\rightarrow	$Det N1$		
		$\langle Det sem \rangle$	=	$\langle NP sem status \rangle$
		$\langle NP syn agr \rangle$	=	$\langle NP sem spec agr \rangle$
		$\langle N1 syn agr \rangle$	=	$\langle NP syn agr \rangle$
		$\langle Det syn agr \rangle$	=	$\langle N1 syn agr \rangle$
		$\langle N1 sem \rangle$	=	$\langle NP sem spec desc \rangle$
$N1$	\rightarrow	N		
		$\langle N sem \rangle$	=	$\langle N1 sem head \rangle$
$N1_1$	\rightarrow	$AP N1_2$		
		$\langle AP sem \rangle$	=	$\langle N1_1 sem mod \rangle$
		$\langle N1_2 sem head \rangle$	=	$\langle N1_1 sem head \rangle$
NP_1	\rightarrow	$NP_2 N1[+of]$		
		$\langle NP_2 sem \rangle$	=	$\langle NP_1 sem spec desc spec_1 \rangle$
		$\langle N1 sem \rangle$	=	$\langle NP_1 sem spec desc spec_2 \rangle$
		$\langle N1 syn agr \rangle$	=	$\langle NP_1 sem spec agr \rangle$
		$\langle NP_2 sem status \rangle$	=	$\langle NP_1 sem status \rangle$
NP_1	\rightarrow	$NP_2 PP[+of]$		
		$\langle NP_2 sem status \rangle$	=	$\langle NP_1 sem status \rangle$
		$\langle NP_2 sem \rangle$	=	$\langle NP_1 sem spec desc spec \rangle$
		$\langle PP sem \rangle$	=	$\langle NP_1 sem spec desc set \rangle$

Figure 9: A fragment of the noun phrase grammar

SRI International, Menlo Park, Ca., July, 1977.

Grosz, Barbara J., Joshi, Aravind K. and Weinstein, Scott (1983) Providing a Unified Account of Definite Noun Phrases in Discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, Massachusetts Institute of Technology, Cambridge, Mass., 15-17 June, 1983, pp44-49.

Grosz, Barbara J. and Sidner, Candace L. (1986) Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12, 175-204.

Hobbs, Jerry R. (1978) Resolving Pronoun References. *Lingua*, 44, 311-338.

Kamp, Hans (1975) Two Theories about Adjectives. In Keenan, E. L. (ed.) *Formal Semantics of Natural Language: Papers from a colloquium sponsored by King's College Research Centre, Cambridge*, pp123-155. Cambridge: Cambridge University Press.

Karttunen, Lauri (1986) D-PATR: A Development Environment for Unification-Based Grammars. In

Proceedings of the 11th International Conference on Computational Linguistics, Bonn, 25-29 August, 1986, pp74-80.

Keenan, Edward L. and Faltz, Leonard M. (1978) *Logical Types for Natural Language*. UCLA Occasional Papers in Linguistics, No. 3.

McDonald, David D. (1980) Natural Language Generation as a Process of Decision-Making under Constraints. PhD Thesis, Department of Computer Science and Electrical Engineering, MIT.

Massam, Diane and Roberge, Yves (1989) Recipe Context Null Objects in English. *Linguistic Inquiry*, 20, 134-139.

Shieber, Stuart M. (1986) *An Introduction to Unification-based Approaches to Grammar*. Chicago, Illinois: The University of Chicago Press.

Webber, Bonnie Lynn (1979) *A Formal Approach to Discourse Anaphora*. London: Garland Publishing.