

# The semi-generative lexicon: limits on lexical productivity

Ann Copestake

University of Cambridge, Computer Laboratory  
New Museums Site, Pembroke Street, Cambridge, UK

Ann.Copestake@cl.cam.ac.uk

## Abstract

This paper provides an overview of several different classes of lexical semi-productivity and discusses a general approach to constraining generative devices.

## 1 Introduction

The counterpoint to any discussion of generative devices in the lexicon should be an appraisal of the counterexamples and the limitations to applicability. This is essential for several reasons. Firstly, the limits on generative processes must be investigated in order to get a better understanding of the theoretical mechanisms underlying generativity. In particular, if the concept of pragmatics as distinct from semantics has any theoretical consequences, purely pragmatic accounts of meaning generation can only be appropriate where there are no conventional, language-specific constraints on the process. Secondly, we must provide formal accounts of generativity which can allow for any exceptions which may be found. This leads to an investigation of devices such as defaults and probabilities. Finally, processing systems that aim for precision of interpretation or idiomatic generation must not overgenerate. Analysis systems that postulate massive ambiguity are of little use. One reason for the rather limited uptake of generative lexical devices in practical natural language processing systems has been the lack of techniques for control of their effects. It is much better in practice to ignore productivity and lose a small proportion of examples than to allow it and be able to process nothing!

In this paper I will go through a number of cases where there appear to be conventional constraints on generative processes. Although much of the data is quite well-known, I think it is useful to consider these phenomena together, because this sheds light on some interesting commonalities and distinctions. I will also describe a possible formal approach. This discussion follows earlier work, in particular the account of productivity in lexical rules proposed in Briscoe and Copestake (1999), but in this paper I consider whether this sort of approach can be used in cases where lexical rules aren't applicable.

## 2 Semi-productivity in alternations and sense extension

I take as a starting point the assumption that lexical rules, implemented for instance within a typed feature structure logic, can be used to give an account of derivational morphology, verb alternations and more general sense extensions, such as the count-to-mass conversion exemplified by the animal and meat senses of *rabbit*, *lamb* etc. I am interested in accounts of lexical rules which allow them to be applied productively to previously unseen lexical items. For instance, *fax* and *email* readily undergo the dative alternation (*fax me your resume/fax your resume to me*). Similarly, it is quite clear that the mass/meat usage is possible even for animals which have not generally been eaten: e.g. *crocodile* and *ostrich* may occur as mass terms. It seems desirable on theoretical grounds to allow such uses to be generated even though they may not have been seen before by most native speakers. Similarly, lexical rules provide a

way of automatically allowing for some of the unseen usages of words in a computational implementation.

However, as discussed in detail in Briscoe and Copestake (1999) (henceforth B&C), while there have been many attempts to define narrow classes within which alternations such as dative are fully productive (e.g., Pinker, 1989), it appears that even though the semantic criteria invoked may be very subtle (and difficult to test or motivate independently), exceptions always remain. For instance, *design* and *create* should be in the same class, but have different acceptability with dative in British English.

- (1) John designed / \*created them a bridge

B&C argue that semi-productivity of verb alternations may be accounted for by assuming rules that are sensitive to both type and token frequency effects. Bauer (1983:71f), in a discussion of derivational morphology, argues that the greater ‘item-familiarity’ of lexical items allows judgments of relative novelty / conventionality to be built up in a way that is not possible at the sentential level because there are simply too many possibilities for the frequency of particular combinations to be assessed. The argument applies just as well to alternation and sense extension (for detailed argumentation, see B&C and also Goldberg (1995)).

## 2.1 A probabilistic account of semi-productivity

The specific proposal made in B&C is to define probabilities for particular lexical entries (by which we mean structures which are either stipulated directly or derived by rule) primarily based on their observed frequencies, as shown in Figure 1. This is a standard approach, but the tie up with lexical rule productivity comes in the estimation of unseen uses. We assumed that rule productivity is defined as the ratio of possible input entries and attested output entries as shown in Figure 2 (see Aronoff, 1976). The unseen probability mass for a word form is defined in terms

of the rules which could potentially apply to it, and the relative probabilities depend on the productivity measurement for the rule, as shown in Figure 3. This will yield revised ratios for each given word which can then be normalized to probabilities.

The assumption we make about processing is simply that speakers generally choose high-frequency forms to realize particular meanings and that users choose high-frequency senses when faced with ambiguity. Under this assumption, speakers never consider infrequent senses (derived or otherwise) unless forced to (by syntax, semantics or pragmatics). Speakers will not generally use unattested forms unless they are licensed by a highly productive rule. Obviously there are exceptions. In some genres there is a value to more creative use of language which presumably causes speakers to deliberately use less likely senses. In spoken language, unattested forms may occur when utterance planning is imperfect. But in general, this principle seems rational for communication — indeed it is entirely compatible with Grice’s Maxim of Manner.

Under this account, speakers do not normally generate utterances such as (2) because there are alternative ways of conveying the same (or a similar enough) meaning which are more probable:

- (2) \* John created them a bridge

On the other hand, even though *crocodile* meaning meat might be unattested in a particular speaker’s experience, there is no attested lexical form conveying the same meaning. The speaker always has the option of choosing a more long-winded expression, but there is a preference for brevity. Note also that *crocodile meat* will likely also be unattested and it is reasonable to assume that speakers have judgments of the conventionality of compound nouns, see below.

Blocking can be treated as a special case of this principle: if speakers use higher frequency forms to convey a given meaning, an extended meaning will not become conventionalized if a common synonym exists. This

$$\text{Prob}(\text{lexical-entry} \mid \text{word-form}) = \frac{\text{freq}(\text{lexical-entry with word-form})}{\text{freq}(\text{word-form})}$$

Figure 1: Lexical entry probabilities, from Briscoe and Copestake, 1999

$$\text{Prod}(\text{lexical-rule}) = \frac{M}{N}$$

(where  $N$  is the number of attested lexical entries which match the lexical rule input and  $M$  is the number of attested output entries).

Figure 2: Lexical rule productivity

$$\text{Unseen-pr-mass}(\text{word-form}) = \frac{\text{number-of-unattested-entries}(\text{word-form})}{\text{freq}(\text{word-form}) + \text{number-of-unattested-entries}(\text{word-form})}$$

$$\text{Est-freq}(\text{lex-entry}_i \text{ with word-form}_j) = \text{Unseen-pr-mass}(\text{word-form}_j) \times \frac{\text{Prod}(lr_i)}{\sum \text{Prod}(lr_1), \dots, \text{Prod}(lr_n)}$$

(where  $lr_1 \dots lr_n$  are the  $n$  lexical rules needed to derive the  $n$  unattested entries for word-form $_j$ )

Figure 3: Estimated frequencies for unattested entries

means that we do not have to stipulate a separate blocking principle in interpretation, since the blocked senses will not be attested or will have a very low frequency. And in generation, we assume that higher probability forms are preferred as a way of conveying a given meaning. Thus *beef* blocks *cow*, meaning the meat, for instance, though as discussed in B&C, blocking is not absolute.

The probabilistic approach to semi-productivity can be seen as a matter of performance rather than competence and has to be formalized separately from the symbolic grammar. Nevertheless, we would argue that such an encoding is a necessary part of any account of lexical generativity, though the specific details of the account may well need to be refined (see the discussion of productivity measurements in B&C, for instance). The exact boundaries of the division between the symbolic and probabilistic components are difficult to determine, because it makes sense to encode some hard constraints on rule applicability. Furthermore, rule productivity might be established for well-defined

semantic subclasses as well as for the rule overall. But in the rest of this paper, I will simply assume that some account based on probabilities and rule productivity is plausible, and discuss how it might apply to other lexical and semi-lexical processes.

## 2.2 Compound nouns

English compound noun formation is at the boundary between lexical and syntactic processes. Copestake and Lascarides (1997) propose an account of compound nouns which is based on the approach to semi-productivity in Briscoe and Copestake (1999). In that paper, we reviewed the evidence for limitations on productivity of the noun-noun compound rule in English, observing, in particular, the lack of direct translations for some German compounds, the phenomenon of ‘possessive’ compounds (e.g., *blacksmith’s hammer*, \**blacksmith hammer*) and the different patterns of stress in compounds. We argued that to account for these effects, and for the apparent existence of conventional meanings, it was necessary to assume a range of rela-

tively fine-grained compound schemata. Although nonce compounds which do not fit into these schemata sometimes occur, they can only do so within a rich discourse context. Compound schemata vary in productivity and we adapted the productivity measurement described above so that it was applicable to compounds.

This proposal is an alternative to accounts of English compound formation as fully productive, which do not allow for some compounds being impossible (with a particular interpretation). It is also preferable to fully lexicalized accounts, which don't allow for novel compounds. We argued that compounds generally have default interpretations based either on their normal meanings (for previously attested compounds) or on meanings that are given by productive compound schemata. While novel compounds which do not fit such patterns are possible, they are rare, and require a rich discourse context for their interpretation, which is not the case for most compounds.

### 2.3 Ham sandwiches

The classic *ham sandwich* examples (Nunberg, 1978) involve non-conventional extended uses which are possible in suitably marked contexts. For instance, in (3), said by someone working in a restaurant, *the ham sandwich* has to mean something like *person who ordered a ham sandwich*:

- (3) The ham sandwich is waiting for his check.

I assume that such examples may be generated by a very broad lexical rule, for instance one that converts nouns denoting physical objects to people associated with that object (there are conventional cases of such sense extensions, such as referring to musicians by instrument nouns). Although such a rule is very broad, I do not assume that any noun can be extended to denote anything: for instance, even in a marked context, using a description of a person to denote an object associated with that person seems to be impossible.

A rule such as physical-object-to-person will have a very low productivity according to the formula above, at least for any normal corpus. This is reasonable, since by definition these are non-conventionalized examples. Interpretation is only possible in a very constrained discourse context, and in actual use there is probably a considerable potential for misunderstanding. Non-conventionalized examples are not infrequently found in newspaper and magazine articles, but presumably generally occur where novelty is valued and precise meanings aren't so important.

However, in a subgenre, ham sandwich examples may become conventionalized, in that the productivity measure will go up as speakers produce new examples. The use of novel extensions in texts such as that cited at length by Clark and Clark (1979) are presumably to some extent deliberate word play. However, if we assume that hearers are sensitive to some quite narrow semantic class implicit in the context (e.g., menu items in restaurants, household appliances or whatever), the productivity measurement for that class would be appreciably increased, even after a very small number of examples. It is clear that this is not all that is going on, but at least in general outline the B&C approach seems consistent with the possibility of non-conventionalized sense extension and at least provides an indication of how conventionalization may occur.

### 3 Non-branching syntax rules

The examples below show a number of cases where phrases which look as though they are NPs act as adverbial modifiers (see also Ostler and Atkins, 1992).

- (4) a I'll meet you next week.  
b We meet every September.  
c I'll meet you Tuesday.  
(\* for British English)  
d \* I'll meet you September.

What seems to be going on here is that some temporally-denoting NPs can be used as though they were PPs. One way of achieving the desired effect is to have a rule that converts temporal NPs to PP-like phrases.

Although such a rule cannot be truly lexical, because it has to be able to apply to phrases with determiners, like *every September*, it nevertheless behaves somewhat like a lexical rule, especially in that it has a very specific range of inputs and has an idiosyncratic semantic effect. There is quite clear dialect variation, as shown by the differences in acceptability of (4c) between (most dialects of) British and American English.

In formal or implementational terms, such rules do not present any great problems for a symbolic component expressed in a typed feature structure framework. In fact the LinGO grammar (Copestake and Flickinger, 2000) implements such an approach and covers a wide range of temporal expressions. But doing this requires a rather detailed semantic hierarchy of temporal nouns, which has to capture distinctions such as day of week versus day of month and allow for the effects of determiners such as *every*. It remains to be seen whether this can be done precisely enough to capture all possible cases of modification without overgenerating and without ending up with an absurdly fine-grained semantic specification. What is most worrying is the possibility that as more and more such rules are encoded, the symbolic grammar becomes overcomplex and even more difficult to maintain.

The alternative approach, along the lines of that proposed for lexical rules, is to let the symbolic rule overgenerate and control its application via probabilities. This however runs into difficulties because we are dealing with phrases rather than words, and these may be arbitrarily complex.

- (5) We'll meet every fourth Friday  
that doesn't fall before a holiday  
weekend.

However, even though it is the entire phrase which is acting as a modifier, the distribution we are interested in concerns the core temporal NP (e.g., *Friday*) and the specifier (and a few other words including *next* and *last*). Thus it may be reasonable to assume that speakers have some degree of item-familiarity

based on components of the phrase. One way of partially testing this would be to see whether we could derive appropriate probabilistic constraints automatically from a realistically sized corpus.

If a probabilistic component can be used, it might be exploited in an account of dialect specificity. That is, while for an American English speaker, *Tuesday* has a reasonably high probability of occurring as a modifier, the probability would be much smaller for a British English speaker.

#### 4 Logical metonymy

Logical metonymy, discussed in Pustejovsky (1995) among other places, also shows interesting restrictions, although the data is not as clearcut as with morphological processes. While (6a) below has both the interpretations *reading the book* and *writing the book*, putatively corresponding to the telic and agentive roles respectively, (6b) apparently only has the reading *building the tunnel* (i.e., the agentive meaning). No interpretation exists corresponding to *using the tunnel*, such as *driving through the tunnel*.

- (6) a Kim began the book.  
b Kim began the tunnel.

As far as I am aware, restrictions of this type were first noted by Godard and Jayez (1993).

What makes this interesting is that (7) is perfect.

- (7) Kim began driving through the  
tunnel.

It is also worth noting that even very marked contexts do not seem to make the telic interpretation better. For example, (8b) is not a possible continuation to (8a):

- (8) a The drive to the Alps had been  
long and tiring, and Kim was  
prone to claustrophobia.  
b \*Therefore it was with considerable  
trepidation that Kim  
began the first tunnel.

However note that *tunnel* and similar nouns are possible with *after* and *enjoy*. The follow-

ing are plausible after (8a):

- (9) a But after the first tunnel, Kim  
felt much happier.  
b But much to his surprise, Kim  
enjoyed the first tunnel.

Godard and Jayez (1993) suggest that the constraint is that the telic interpretation is only possible with *begin* when the additional event involves consumption, which (somewhat implausibly) has to be assumed to include reading. Pustejovsky and Bouillon (1995) discuss possible constraints involving aspect and control properties. But Verspoor (1997) demonstrates the inadequacies of both these accounts. Furthermore, the corpus data she describes show that the overwhelming preponderance of cases of logical metonymy with *begin* and a putative telic interpretation involve a very limited class of physical object/substance nouns, especially nouns denoting foodstuffs, drinks and books. The distribution is relatively similar for *finish*, which shows a greater frequency of metonymy, although metonymies with physical object/substance noun phrases form a very small proportion of the uses of both verbs.<sup>1</sup>

It might be possible to attempt an account where the metonymic process, however it is encoded, applies only to a finely specified semantic class. But the data suggests that we would at best end up with a disjunctively specified class, or equivalently, with a set of subcases each concerning a very finely specified class. This has obvious analogies with lexical alternations. Logical metonymy may be sufficiently infrequent for the item familiarity story to be plausible here: although

---

<sup>1</sup>Verspoor also notes that nouns like *story* and *song* occur with *begin* and *finish* in the sense of *tell/perform* but this might alternatively be classified as agentive. The data is also complicated because Verspoor excludes eventive nouns in her definition of metonymy, and while this distinction may be justified, it is a little hard to make precise. Her comparison with other verbs also raises interesting issues. For my current purposes, however, all that matters are the limitations on the use of *begin* and *finish* with nouns denoting physical objects or substances, since this is not predicted on a fully productive account of metonymy.

it is necessary to consider the interaction of two words which don't generally occur next to each other in the string, the existence of verb-noun collocations (discussed below) suggest that this is plausible. Of course this means assuming that the probabilities concern the interaction of the verb and the head of the noun phrase rather than the phrase as a whole.

## 5 Collocations

Collocations have been generally neglected within the Generative Lexicon literature, although they are a major focus of attention within Meaning-Text Theory (e.g., Mel'čuk and Polguère (1987)) and there has been a considerable amount of work on them in computational linguistics. I will use collocation to mean two or more lexical items occurring together in the same sentence more frequently than would be expected, given a fully adequate symbolic grammar and taking into account world knowledge. For instance, *shake* and *fist* are collocates, while *buy* and *house* are probably not. Although the latter pair cooccur more often than would be expected by chance, this may well be predictable given the role of house purchase in our culture. This definition does not lend itself to a direct test, but there is potential for investigating it via WordNet synonym sets, for instance (see also Pearce, 2001).

The converse of a collocation is an *anti-collocation*, where lexical items cooccur less frequently than would be expected and a phrase might be regarded as odd by a native speaker.<sup>2</sup> For instance, Cruse (1986:281) observes that *impeccable behaviour*, *impeccable performance* and *flawless performance* are all natural but that *flawless behaviour* is slightly odd. In at least some cases, the intuition is that the phrase is blocked by a collocation: for instance, *concentrated tea* is odd perhaps because of the naturalness of *strong tea*. Meaning-Text Theory hardwires such effects via functions such as Magni (Magni-

---

<sup>2</sup>Although this concept has been discussed in the literature, there doesn't appear to be any standard terminology: the term anti-collocation is taken from Pearce (2001).

tude). However, the phenomenon of anti-collocation lends some support to using an approach based on probabilities such as that in B&C instead. If a collocation such as *strong tea* is frequent enough that a speaker is aware of it as a collocation, then a different phrase with a very similar meaning would be predicted to be dispreferred. However the blocking effect might be expected to be weaker with phrases than with lexical items because of their lower item familiarity.

It is often difficult to draw a firm line between collocation and a semantic explanation. For instance, *heavy smoker* and *heavy drinker* are normally described as collocations, but it could be argued that there is simply a sense of *heavy* which applies specifically to consumption. This is partly borne out by the acceptability of *heavy use* and *heavy consumption*, but it would be necessary to refine the concept to account for the unacceptability of *heavy eater*. Furthermore, other senses of *heavy* would be needed to account for *heavy weather*, *heavy sea*, *heavy breathing* and so on. In some cases there does not seem to be a realistic semantic hypothesis at all: consider the problem of accounting for the distribution of *bake* versus *roast* given the contrast between *pork* and *ham* which does not depend on any consistent difference in cooking method (*baked ham*, *?roast ham*, *roast pork*, *?baked pork*).

There may well be cases where there is considerable discrepancy between the perceptions of native speakers, especially for less frequent words. For instance, *rancid* occurs with a wide range of nouns, but it is clear to all the speakers I have asked that its core meaning is something to do with offness in food. Different speakers have different intuitions: for instance that it refers to dairy products, to fats and oils or to fatty food. The acceptability of *rancid meat* thus varies. However there is a technical definition, which involves off flavours or smells caused by fat oxidation. It is therefore impossible to say whether *rancid butter* (which is a core use for most of the speakers I have asked) is a collocation or whether its frequency is predictable (based on a limited adjectival meaning and world knowl-

edge about food), because individual speakers differ. But this discrepancy is unlikely to be noticed outside a technical context. It is worth noticing that most people will not have been exposed to a very large number of instances of the word *rancid*: it occurs only 77 times in the British National Corpus, and many of these examples must presumably be seen as non-core (e.g., *rancid T-shirt*, *rancid voice*, *rancid first quarter of the century*).

I would argue that, as with verb alternations and logical metonymy, attempting a fine-grained semantic account of acceptability for all cases of possible collocation does not seem plausible. It is interesting that non-linguists are more able to provide rationalizations of word use restrictions than of alternations, and it is certainly possible that some speakers have a specific meaning hypothesis for some potential collocations, but it is very unlikely that all speakers have such hypotheses for all of them. It seems more likely that there is some degree of conventionalization based on frequency, and that this leads to a certain amount of semi-productivity even in phrase formation. From a computational perspective, attempting to derive fine-grained meanings for an adjective like *heavy* is unattractive. It is much more feasible to assume a broad meaning and to restrict its cooccurrences via a frequency-based mechanism based on attested combinations. This would allow an estimation of likelihood for unseens based on a similar technique to that described in B&C: that is, treating productivity on the basis of attested combinations within word classes. It is possible to regard this as a way of allowing meaning to be encoded implicitly via knowledge of the possibilities for semantic combination.

## 6 Conclusion

I have tried to show the pervasiveness of semi-productivity in lexical processes and to argue that a reasonable way to account for it is to limit a fully productive symbolic account via a frequency-based mechanism. The line between cases which should be totally ruled out and those which should be dispreferred is not

totally clear and perhaps never can be clear, since there is unlikely to be perfect agreement between speakers. But at the limit there is no observable difference between assuming that a particular form or phrase has an infinitesimal probability and ruling it out completely.

This discussion of the limitations of generativity is not intended as an attack on Generative Lexicon theory. In fact, quite the opposite is true. The only viable alternatives to the assumption of generative devices in the lexicon are approaches that leave these processes to pragmatics. The work of Hobbs (e.g., Hobbs (1979), Hobbs (1985), Hobbs et al (1993)) is particularly important in providing an account of a wide range of phenomena. But pragmatic approaches have real problems in dealing with conventional limitations to generativity, especially where these are language-specific or dialect-specific. Any approach which assumes that operations like metonymy or logical metonymy are carried out at the level of logical form can only allow for conventional restrictions via operations that affect the logical form, and this just does not seem a reasonable way of accounting for most of the data described above. So, perversely, limitations on generativity provide the best arguments for a (semi-)generative lexicon.

## References

- Aronoff, Mark (1976) *Word Formation in Generative Grammar*, Linguistic Inquiry Monograph 1. MIT Press, Cambridge, Mass.
- Bauer, Laurie (1983) *English Word-Formation*, Cambridge University Press, Cambridge, England.
- Briscoe, Edward J. and Ann Copestake (1999) 'Lexical rules in constraint-based grammars', *Computational Linguistics*, **25:4**, 487–526.
- Clark, Eve V. and Herbert H. Clark (1979) 'When nouns surface as verbs', *Language*, **55**, 767–811.
- Copestake, Ann and Dan Flickinger (2000) 'An open-source grammar development environment and broad-coverage English grammar using HPSG', *Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece.
- Copestake, Ann and Alex Lascarides (1997) 'Integrating symbolic and statistical representations: the lexicon-pragmatics interface', *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL-EACL 97)*, Madrid, pp. 136–143.
- Cruse, D. A. (1986) *Lexical semantics*, Cambridge University Press, Cambridge, England.
- Godard, Danièle and Jacques Jayez (1993) 'Towards a proper treatment of coercion phenomena', *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics (EACL-93)*, Utrecht, The Netherlands, pp. 168–177.
- Goldberg, Adele (1995) *Constructions*, Chicago University Press, Chicago, IL.
- Hobbs, Jerry (1979) 'Coherence and Coreference', *Cognitive Science*, **3**, 67–90.
- Hobbs, Jerry (1985) 'On the Coherence and Structure of Discourse', Report No. CSLI-85-37, Center for the Study of Language and Information.
- Hobbs, Jerry, Mark Stickel, Doug Appelt and Paul Martin (1993) 'Interpretation as Abduction', *Artificial Intelligence*, **63.1**, 69–142.
- Mel'čuk I and Polguère M (1987) 'A formal lexicon in Meaning-Text Theory (or how to do lexica with words)', *Computational Linguistics*, **13:3–4**, 261–275.
- Nunberg, Geoffrey D. (1978) 'The pragmatics of reference', Doctoral Dissertation, CUNY Graduate Center, reproduced by the Indiana University Linguistics Club.
- Ostler, Nicholas and B.T.S. Atkins (1992) 'Predictable meaning shift: some linguistic properties of lexical implication rules' in J. Pustejovsky and S. Bergler (ed.), *Lexical Semantics and Knowledge Representation. Proceedings of the*

- first SIGLEX Workshop, Berkeley, CA*, Springer-Verlag, Berlin, pp. 87–100.
- Pearce, Darren (2001) ‘Using conceptual similarity for collocation extraction’, *Proceedings of the Fourth annual CLUK colloquium*, Sheffield, UK.
- Pinker, Steven (1989) *Learnability and Cognition: The Acquisition of Argument Structure*, MIT Press, Cambridge, MA.
- Pustejovsky, James (1995) *The Generative Lexicon*, MIT Press, Cambridge, MA.
- Pustejovsky, James and Pierrette Bouillon (1995) ‘Aspectual coercion and logical polysemy’, *Journal of semantics*, **12:2**, 133–162.
- Verspoor, Cornelia M. (1997) ‘Contextually-dependent lexical semantics’, PhD thesis, Centre for Cognitive Science, University of Edinburgh.