Chapter 12
Pragmatics in the (English) lexicon[1]

Keith Allan

1. Introduction

In this chapter I shall discuss only the lexicon of English, but the general principles seem to apply to many, if not all, other languages even though the minutiae do not. By "lexicon" I mean a rational model of the mental lexicon or dictionary. Although the way a lexicon is organized depends on what it is designed to do, it is minimally necessary for it to have formal (phonological and graphological), morphosyntactic (lexical and morphological categorization) and semantic specifications. Relations are networked such that formal specifications are (bi-directionally) directly linked to morphosyntactic specifications that are directly linked to semantic specifications – which, for the moment, subsumes pragmatic specifications. A lexicon must be accessible from three directions: form, morphosyntax, and meaning; none of which is intrinsically prior. Each of these three access points is, additionally, bi-directionally connected with an encyclopaedia.Haiman 1980: 331 claimed "Dictionaries *are* encyclopaedias" and certainly many desk-top dictionaries contain extensive encyclopaedic information (e.g. Hanks (ed.) 1979; Kernfeld 1994; Pearsall (ed.) 1998). The position taken here is that a lexicon is a bin for storing listemes[2], language expressions whose meaning is (normally) not determinable from the meanings (if any) of their constituent forms and which, therefore, a language user must memorize as a combination of form, certain morphosyntactic properties, and meaning. An encyclopaedia is a structured data-base containing exhaustive information on many (perhaps all) branches of knowledge. It therefore

1. My thanks to Kasia Jaszczolt for making me clarify bits of this chapter. Kasia is not to blame for remaining infelicities; indeed, she heartily disapproves some of my claims.

2. The term *listeme* is from Di Sciullo and Williams 1987. Listemes may consist of a single morpheme (such as PAST TENSE), a lexeme (such as TAKE), a multiword "prefab" (*put up with, shoot the breeze, doesn't amount to a hill of beans*, see §9) and perhaps potentially productive stems such as –JUVENATE (see Allan 2001). Listemes are (apparently) what Stubbs 2001calls "lemmas" and Wray 2008 calls "morpheme equivalent units".

seems more logical that the lexicon forms part of an encyclopaedia than vice versa, but the actual relationship does not significantly affect this article. I assume that encyclopaedic information is typically, if not uniquely, pragmatic.

A lexicon is a bin for storing listemes for use by language speakers in any and all contexts. This is not to deny that new listemes are occasionally created, but the coining of a new listeme is a rare event and the resources of a lexicon are normally adequate for all contexts that a speaker faces. Consequently the meanings of listemes are expected to be adapted by semantic extension or narrowing both concretely and figuratively by speakers in utilising them and hearers in interpreting them. Such lexical adjustment can be illustrated by the various meanings of the related listemes *cut* in **Error! Reference source not found.**.

(1) cut grass, cut hair, cut steel, cut the thread, cut the cards, cut your losses, cut out the middle man, cut the ties, to cut and run, cut the cackle, cut a class, cut someone socially, be a cut above, she's all cut up by the breakdown in her marriage, be cut to the quick, cut through the obfuscation, cut my finger, cut the tyres, cut the cake, cut a disk, a railway cutting, cut through the back lane, cut a [fine] figure

Most, if not all, of these seem to derive from a basic notion of severing, interpreted in various ways according to what is severed and/or the manner of severing (this could even apply to *cut a figure*). Similarly, it is well-known that a colour term may extend to shades very far from the focal colour (Berlin and Kay 1969; MacLaury 1997) as selected from, say, the Munsell Color Array; we can attribute this to the elasticity that language needs to have in order that it can usefully be applied to the world around us. In certain domains and in certain formulaic expressions colour terms are used of hues vastly distant from the focal colour. Take the domain of human appearance: terms like *white, black, yellow,* and *brown* have all been used to characterize the skin pigmentation of people of different races, often dysphemistically. These colour terms are descriptively appropriate not so much in relation to the focal colours as in relation to each other: a *white person* is typically paler than the others and a *black person* darker; a *yellow person* is typically yellower than the others. The peoples of south east Asia and Austronesia are often referred to as *brown*, despite the fact that peoples labelled

black are often of similar brown skin colour. So *brown*, too, functions by contrast with *white, black* and *yellow* in this domain. In the domain of oenology, *red wine* does have a (usually dark) red tinge but *white wine* is only white by virtue of being paler than red wine; white wine is normally pale yellow or pale green. Clearly what determines the meanings of these particular sets of colour terms is their comparative function: by means of very rough approximation to the focal colour, they distinguish within a semantic field between different species of the kind of entity denoted by the noun they modify.

Pragmatics within the lexicon is largely an addition to the semantic specifications; for instance, it is useful to identify the default meanings and connotations of listemes. Default meanings are those that are applied more frequently by more people and normally with greater certitude than any alternatives. Bauer 1983: 196 proposed a category of "stylistic specifications" to distinguish between *piss, piddle*, and *micturate*, i.e. to reflect the kind of metalinguistic information found in traditional desk-top dictionary tags like 'colloquial', 'slang', 'derogatory', 'medicine', 'zoology'; such metalinguistic information is more encyclopaedic than lexical. So too is etymological information. Pustejovsky 1995: 101 specifies *book* as a "physical object" that "holds" "information" created by someone that "write[s]" it and whose function is to be "read". Certainly, there is a relation between *book*, *write*, and *read* that needs to be accounted for either in the semantic specification or pragmatically – Pustejovsky represents it in terms of a network and networks are also used in frame semantics (Fillmore 1982; 2006; Fillmore and Atkins 1992; FrameNet at http://framenet.icsi.berkeley.edu) and by Vigliocco, Meteyard, Andrews et al. 2009. Category terms like *noun, verb*, *adjective*, and *feminine* are part of the metalanguage, not the object language; but they also appear in the lexicon as expressions in the object language and there needs to be a demonstrable relation from object language to metalanguage (and vice versa). It would seem incontrovertible that encyclopaedic data is called upon to interpret non-literal expressions like *Ella's being a tiger*; likewise, to explain the extension of a proper name like *Hoover* to denote vacuum cleaners and vacuum cleaning or the formation of the verb *bowdlerize* from the proper name *Bowdler*. I assume that, because many proper names are

shared by different name-bearers, there must be a stock of proper names located either partially or wholly in the lexicon, even if they are stored differently in the brain (see §9). The production and interpretation of statements like those in (2)–(3) requires pragmatic input.

(2) Caspar Cazzo is no Pavarotti!

(3) Harry's boss is a bloody little Hitler!

(2) implies that Caspar is not a great singer; we infer this because Pavarotti's salient characteristic was that he was a great singer. (3) is abusive because of the encyclopaedic entry for the name *Hitler* that carries biographical details of a particular name bearer. Such comparisons draw on biodata that are appropriate in an encyclopaedia entry for the person who is the standard for comparison but not appropriate in a lexicon entry; the latter should identify the characteristics of the typical name-bearer, such as that *Aristotle* and *Jim* are normally names for males, but not (contra Frege 1892) the biographical details of any particular name bearer – any more than the dictionary entry for *dog* should be restricted to a whippet or poodle rather than the genus as a whole.

One of the earliest investigations of lexical pragmatics was McCawley 1978, McCawley (correctly) argued that a listeme (such as *pink* or *kill*) and a semantically equivalent paraphrase (such as *pale red* or *cause to die*) are subject to different pragmatic conditions of appropriateness that give rise to different interpretations, which he thought could be captured by general conditions of cooperative behaviour such as Grice's cooperative maxims. He did not tackle the question of whether pragmatics intrudes on lexical entries. Nor do Blutner 1998; 2004; 2009. Blutner discusses pragmatic compositionality, blocking (if a listeme already exists to express a meaning, do not construct another one without good reason to do so[3]), and pragmatic anomaly (recognized as early as Apollonius Dyscolus in *Peri Suntaxeōs* III.149, see Uhlig (ed.) 1883). The closest Blutner comes to pragmatics within the lexicon is discussing the interpretation of certain adjectives and institute-type nouns (Blutner 1998).

3. For discussion of its implementation and exceptions see Allan 2001 and references cited there.

Carston 2002 (Ch.5) then Wilson and Carston 2007discuss lexical narrowing (e.g. *drink* used for 'alcoholic drink'), approximation (e.g. *flat* meaning 'relatively flat') and metaphorical extension (e.g. *bulldozer* used to mean 'forceful person'). They argue that the same interpretive processes as are employed for literal utterances are used for narrowing, broadening, through to approximation and figurative usage in hyperbole and metaphor. Interpretation is triggered by the search for "relevance" constrained by the principle of least effort: "An input is relevant to an individual when it connects with available contextual assumptions to yield positive cognitive effects (e.g. true contextual implications, warranted strengthenings or revisions of existing assumptions)" (Wilson and Carston 2007: 245). Inferences deriving from "explicature", "implicature", and context-based assumptions satisfy the expectation of relevance, which causes the interpretive process to stop at whatever interpretation a hearer judges satisfactory in the context of utterance.

Huang 2009 also deals with lexical narrowing, lexical blocking, and pragmatic anomaly and, in addition, contrastive focus reduplication. But (despite his title "Neo-Gricean pragmatics and the lexicon") he has very little more to say about pragmatics in the lexicon than is found in Blutner or Wilson and Carston.

Copestake and Lascarides 1997 identified the importance of noting in the lexicon the frequency of particular word senses, in a manner very similar to that independently proposed for a broader range of data by Allan 2000; 2001and again in this chapter. Copestake and Lascarides 1997: 140 write "For example, in the BNC [British National Corpus] *diet* has probability of about 0.9 of occurring in the food sense and 0.005 in the legislature sense (the remainder are metaphorical extensions, e.g. *diet of crime*)." In §2 of this chapter I introduce a credibility metric like that of Copestake and Lascarides which applies to (some) nonmonotonic statements within the lexicon. I argue the case for nonmonotonic statements in the lexicon in entries for nouns in §3 and for verbs in §4. In §5 I discuss the pragmatic intrusions into the interpreting of collectives and collectivized nouns. This leads naturally to a consideration in §6 of the entries for animal nouns that may refer to either the animal's meat or its pelt (after Allan 1981; Nunberg and Zaenen 1992); §7 takes up the dictionary entry for

and; §8 discusses the pragmatic component of lexicon entries for sorites terms. §9 looks at the place of "prefabs" or "formulaic expressions" in the lexicon and §10 tackles ways in which connotation might be incorporated into entries for listemes. §11 summarises the chapter.

2. A credibility metric

In some of what follows it will be helpful to use a credibility metric for a proposition. The truth value of a proposition *p* hinges on whether or not *p* is, was or will be the case. What matters to language users is not so much what is in fact true, but what they believe to be true.[4] The credibility of *p* is what is believed with respect to the truth of *p*, or believed is known, or is in fact known of its truthfulness. Because most so-called 'facts' are propositions about phenomena as interpreted by whomever is speaking, we find that so-called 'experts' differ as to what the facts are (for instance, wrt global warming, or what should be done about narcotics, or what is the best linguistic theory). Whether ordinary language users judge a proposition true or false depends partly on its "pragmatic halo" (Lasersohn 1999): in any normal situation *Sue arrived at three o'clock* is treated as true if she arrived close to three o'clock; the slack afforded by the pragmatic halo is restricted by a pragmatic regulator such as *precisely* or *exactly* in *Sue arrived precisely at three o'clock* or *Sue arrived at exactly three o'clock*.[5] Mostly, though, truth or falsity is assigned by the ordinary language user on the basis of how credible the proposition is, and this is reflected in the way that language is produced and understood. There is a credibility metric such as that in Table 12.1, in which complete confidence that a proposition is true rates 1, represented CRED = 1, and complete confidence that a proposition is false rates CRED = 0; indeterminability is midway between these two, CRED = 0.5. Other values lie in between. (□ is the necessity operator, ◇ is the possibility operator, $\underline{\vee}$ symbolizes exclusive disjunction, ¬*p* means "not-*p*".)

4. Religious conflicts make this very obvious.
5. Lasersohn thinks this erases the slack, but I think the slack is only restricted.

Table 12.1. The credibility metric for a proposition

CRED = 1.0	Undoubtedly true:	$\Box p$, I know that *p*
CRED = 0.9	Most probably true:	I am almost certain that *p*
CRED = 0.8	Probably true:	I believe that *p*
CRED = 0.7	Possibly true:	I think *p* is probable
CRED = 0.6	Just possibly true:	I think that perhaps *p*
CRED = 0.5	Indeterminable:	$(\Diamond p \geq 0.5) \veebar (\Diamond \neg p \leq 0.5)$
CRED = 0.4	Just possibly false:	It is not impossible that *p*
CRED = 0.3	Possibly false:	It is not necessarily impossible that *p*
CRED = 0.2	Probably false:	It is (very) unlikely that *p*
CRED = 0.1	Most probably false:	It is almost impossible that *p*
CRED = 0.0	Undoubtedly false:	$\Box \neg p$, I know that $\neg p$

In reality, one level of the metric overlaps an adjacent level so that the cross-over from one level to another is more often than not entirely subjective; levels 0.1, 0.4, 0.6, 0.9 are as much an artifact of the decimal system as they are independently distinct levels in which I have a great deal of confidence. Nonetheless, I am certain that some variant of the credibility metric exists and is justified by the employment of the adverbials (*very*) *probably*, (*very*) *possibly* and *perhaps* in everyday speech. This metric is needed in some lexical entries, as we shall see.

3. Semantic specifications for *bird* and *bull*

Birds are feathered, beaked, and bipedal. Most birds can fly. Applied to an owl this attribute of flight is true; applied to a penguin it is false. Birds are sexed and a normal adult female bird can lay eggs. It is a defining characteristic that members of the female sex carry ova; I'll label this function SXF (which can be glossed 'sexual female'). Where they don't, or the ova are non-viable, the organism can count for our purposes as a gendered female, GENF, but not SXF. Mostly, sexual females are gendered females too; see (4) where → indicates semantic entailment.

(4) MOST(x)[SXF(x) → GENF(x)]

Although we do speak of human eggs, nonetheless the default egg is from an oviparous genus such as a bird, so I'll assume this characteristic ought to be noted in the lexicon.[6] Based on Allan 2001: 252, I propose that the semantic part of the lexicon entry for *bird* be (5), where ∧ symbolizes logical conjunction, +> indicates (defeasible) nonmonotonic inference (NMI), which could perhaps be referred to as an implicature and which is cancelled for species such as emus and penguins.

(5) $$\forall x \left[\begin{array}{l} \text{BIRD}(x) \rightarrow \lambda y[\text{FEATHERED}(y) \wedge \text{BEAKED}(y) \wedge \text{BIPEDAL}(y)](x) \\ \text{BIRD}(x) +> \Diamond\text{FLY}(x),\ \text{CRED} \geq 0.7 \\ \lambda z[\text{BIRD}(z) \wedge \text{SXF}(z) \wedge \text{ADULT}(z)](x) \rightarrow \text{OVIPAROUS}(x) \end{array}\right]$$

The lambda-operator is useful to identify an individual as having a number of properties jointly, e.g. being a member of the set of creatures that are at the same time feathered and beaked and bipedal. In (5) the line BIRD(x) +> ◇FLY(x) identifies that a bird is most probably capable of flight with a credibility rating of 0.7. In the case of a sparrow, the semantic component of the lexicon entry may look like (6); for a penguin, like (7).

(6) $$\forall x \left[\begin{array}{l} \text{SPARROW}(x) \rightarrow \text{PASSERINE}(x) \\ \text{PASSERINE}(x) \rightarrow \lambda y[\text{BIRD}(y) \wedge \Diamond\text{FLY}(y)](x),\ \text{CRED} \geq 0.99 \end{array}\right]$$

(7) $$\forall x \left[\begin{array}{l} \text{PENGUIN}(x) \rightarrow \text{SPHENISCIDA}(x) \\ \text{SPHENISCIDA}(x) \rightarrow \lambda y[\text{BIRD}(y) \wedge \neg\text{FLY}(y)](x),\ \text{CRED} = 1 \end{array}\right]$$

For both (6) and (7) the oviparity of SXF sparrows and penguins is an entailment of their being birds. The credibility of a sparrow being able to fly is estimated at CRED ≥ 0.99 (it might be injured), whereas the credibility of a penguin flying is 0 (its not-flying has a credibility of 1).

The first entry under *bull* in the *Oxford English Dictionary* 1989 is "The male of any bovine animal; most commonly applied to the male of the domestic species (*Bos Taurus*); also of the buffalo, etc." Part of this is more formally stated in (8).

(8) $\forall x[\lambda y[\text{BULL}(y) \wedge \text{ANIMAL}(y)](x) \rightarrow \lambda z[\text{MALE}(z) \wedge \text{BOVINE}(z)](x)]$

6. One reconstruction of the Proto-Indo-European word for EGG is **$h_a\bar{o}(w)iom$* "bird-thing" from **$h_ae(w)ei$-* "bird" (I am grateful to Olav Kuhn for this information).

I will ignore the facts identified in (9).

(9) $\text{MALE}(x) \rightarrow \text{GEN}M(x) +> \text{SX}M(x)$

(8) is inaccurate because the noun *bull* is not restricted in application to bovines; it is also properly used of male elephants, male hippos, male whales, male seals, male alligators, and more. The initial plausibility of (8) is due to the fact that it describes the stereotypical bull. The world in which the English language has developed is such that *bull* is much more likely to denote a bovine than any other species of animal. Peripheral uses of *bull* are examples of semantic extension from bovines to certain other kinds of large animals; consequently they require that the context make it abundantly clear that a bovine is not being referred to. This is often achieved by spelling it out in a construction such as *bull elephant* or *bull whale* which is of greater complexity than the simple noun *bull* used of bovines – a difference motivated by the principle of least effort (Zipf 1949). There is no regular term for "the class of large animals whose males are called 'bulls', females 'cows', and young 'calves'" so in Allan 2001: 273 I coined the term **bozine* to label it.[7] The semantics of English *bull* is given in (10) from which the NMI of bovinity will be cancelled where the animal is contextually specified as giraffid, hippopotamid, proboscid, pinniped, cetacean, or crocodilian.

(10) $\forall_x \left[\begin{array}{l} \lambda y[\text{BULL}(y) \wedge \text{ANIMAL}(y)](x) \rightarrow \lambda z[\text{MALE}(z) \wedge \text{*BOZINE}(z)](x) \\ \lambda y[\text{BULL}(y) \wedge \text{ANIMAL}(y)](x) +> \text{BOVINE}(x), \text{CRED} \geq 0.9 \end{array} \right]$

Once again we see a default interpretation being recorded as a NMI in the lexicon because of the salience of this particular characteristic, viz. bovinity, of the default reference (i.e. the denotatum) for *bull*. (At first sight a salient meaning should be almost the opposite of a default meaning: something that is salient jumps out at you; by contrast a default is the fall-back state when there is no contextual motivation to prefer any other. On a second look, what qualifies a state to become the default is its salience in the absence of any contextual motivation to prefer another.) The credibility of ≥0.9 is based on my intuition. A search of ten

7. The fact that there is no word for *bozines is suggests either that English speakers can function with the vague category 'large animals, like bovines are' or that terms such as *bull elephant* and *cow whale* are learned first and *elephant calf* and *bull whale* can be adduced by analogy.

corpora totalling about 10 million words (the Australian corpus of English; Australian ICE; the Lancaster–Oslo/Bergen corpus of British texts; the London–Lund corpus; the Freiburg corpus of British texts; the Freiburg corpus of American texts; the Brown corpus of American texts; the Wellington corpus of written New Zealand texts; New Zealand ICE; Kenya –East Africa ICE) revealed no applications of *bull* to animals other than bovines, nor indeed were such searches useful in confirming or disconfirming any of the other credibility ratings in this chapter.

In this section I have shown that a lexicon entry can be constructed to indicate the necessary components of meaning for the entry and also the most probable additional components of meaning that obtain for most occasions of use but which may be cancelled as a function of contextual constraints. These can be seen as prototype effects that, for instance, help distinguish *cup* from *mug* and *bowl* (see Labov 1978). Traditional Arab and Turkish coffee cups are small bowls with no handle, very similar in configuration to Chinese porcelain tea-cups. The typical Western tea-cup or coffee cup has a handle and is accompanied by a saucer. All these types of cup are bowl-like in shape though they are smaller, usually have higher sides, and serve a different function than most bowls. Cups are intended to be put to the lips to convey liquid to the mouth whereas liquid in food bowls is spooned into the mouth; otherwise a bowl is used for food preparation. These kinds of conditions (that distinguish *cup* from *mug* and *bowl*) are encyclopaedic and pragmatic rather than purely semantic.

For each lexicon entry the semantic identity of the listeme is presented as a meaning postulate, cf. (10); for instance, the noun *bull* is semantically represented by the predicate BULL ranging over a variable for the entity denoted. Predicates like BULL, ANIMAL, MALE, and BOVINE are not decomposed into semantic primitives but give rise to certain inferences some of which are necessary semantic entailments, others are probabilistic nonmonotonic inferences. Similar conditions apply to the verb *climb*, as we see in §4.

4. Climbing

Jackendoff 1985 identified some interesting characteristics of the verb *climb*. From (11) we understand that Jim climbed up the mountain – contrast (11) with (12). We also understand that he used his legs and feet – contrast (11) and (12) with (13).

(11) Jim climbed the mountain.

(12) Jim climbed down the mountain.

(13) Jim climbed (down) the mountain on his hands and knees.

Snakes, airplanes, and ambient temperature lack legs and feet they can use when climbing (which is presumably a metaphorical extension with these actors), and they can't *climb down*, some other verb must be employed.

(14) The snake climbed { the tree. / ?? down the tree.

(15) The airplane climbed { to its cruising altitude. / ?? down to land.

(16) The temperature climbed { to 42. / ?? down to minus 10.

In (17) the lexicon entry captures the fact that the default interpretation of *climb* presumes both upward movement, symbolized by ↑[8] **and** the use of feet (and therefore legs, too).

(17) $\forall x \left[\begin{array}{l} \text{CLIMB}(x) \rightarrow \lambda y[\text{GO}(y)\_\uparrow \vee \text{USE\_FEET}(y)[\text{CAUSE}(y)[\text{MOVE}(y)\_\uparrow]]](x) \\ \text{CLIMB}(x) +\!> \lambda y[\text{GO}(y)\_\uparrow \wedge \text{USE\_FEET}(y)[\text{CAUSE}(y)[\text{MOVE}(y)\_\uparrow]]](x),\ \text{CRED} \approx 0.7 \end{array} \right]$

NMI apply not just to nouns and verbs but potentially in any lexicon entry.

5. Collectives and collectivizing

Allan 1976; 2001 discuss the semantics of collective nouns such as *admiralty, aristocracy, army, assembly, association, audience, board, class, clergy, committee, crowd, flock, government* and collectivized nouns such as those italicized in (18)–(19).

8. This 90º from the horizontal is the prototype for "upward", but any angle greater than 0 and less than 180º is upward.

(18) These three *elephant* my great-grandfather shot in 1920 were good tuskers, such as you never see today.

(19) Four silver *birch* stand sentinel over the driveway entrance.

A definition of collectivizing will be given shortly, but let's begin with familiar collectives.

Collective nouns allow reference to be made to either the set (collection) as a whole or to the set members. In many dialects of English (but not all) the different interpretations are indicated by NP-external number registration; consider (20).[9]

(20) The herd $\left\{ \begin{matrix} \text{is} \\ \text{are} \end{matrix} \right\}$ getting restless and $\left\{ \begin{matrix} \text{it is} \\ \text{they are} \end{matrix} \right\}$ beginning to move away.

Whereas singular NP-external number registration indicates that the set as a holistic unit is being referred to, cf. (21), the plural indicates that the set members are being referred to, (22). In these and later examples, X and Y are (possibly null) variables for NP constituents; NP_{SG} is a singular NP, and NP_{PL} is plural; x, y, z are sets, either unit sets (individuals)[10] or multimember sets, so one should understand from (21) and (22) that $\forall x[\exists y[y \subseteq x]]$.

(21) $\forall x[NP_{SG}[X\ N_{HEAD}[\lambda y[\text{MANY}(y) \wedge \text{COLLOCATED}(y)](x)]\ Y]$
$\rightarrow \text{COMBINED\_MEMBERSHIP}(x)]$

(22) $\forall x[NP_{PL}[X\ N_{HEAD}[\lambda y[\text{MANY}(y) \wedge \text{COLLOCATED}(y)](x)]\ Y]$
$\rightarrow \text{CONSTITUENT\_MEMBERS}(x)]$

Thus, (23) identifies the composition of the committee, while (24) identifies dissension among the membership of the committee.

(23) The committee $\left\{ \begin{matrix} \text{is} \\ \text{?*are} \end{matrix} \right\}$ composed of many notable scholars.

(24) The committee $\left\{ \begin{matrix} \text{?*is} \\ \text{are} \end{matrix} \right\}$ at odds with each other over the new plan.

NPs denoting institutions, e.g. *the company I work for, the BBC, the university* must be singular (NP_{SG} in (27) and (28)) when the institution as a building, location, or single

9. It is assumed here that countability is characteristic of NPs rather than nouns, as argued in Weinreich 1966, McCawley 1975, and Allan 1980.
10. There is no evidence that natural languages distinguish between individuals and unit sets.

constituent body is referred to, as in (25), but can have plural NP-external registration when referring to the people associated with it, (26).

(25) The library $\left\{\begin{matrix}\text{is}\\ \text{?*are}\end{matrix}\right\}$ located in the new civic centre.

(26) The library $\left\{\begin{matrix}\text{charges}\\ \text{charge}\end{matrix}\right\}$ a heavy fine on overdue books.

The facts with respect to such collective nouns are represented in (27)–(29), where N_0 is the form of the noun unmarked for number.

(27) $\forall x \exists z[N_0[\text{LIBRARY}(x)] \rightarrow \lambda y[\text{MANY}(y) \wedge \text{BOOK}(y) \wedge \text{COLLOCATED}(y)](z) \wedge x \supseteq z]$
$+> \exists x[NP_{SG}[X\ N_{0,HEAD}[\text{LIBRARY}(x)]\ Y] \wedge \text{INSTITUTION}(x)]$

(28) $\forall x[NP_{SG}[X\ N_{0,HEAD}[\text{INSTITUTION}(x)]\ Y] \rightarrow \text{CONSTITUENT\_BODY}(x) \vee \text{SITE}(x)]$

(29) $\forall x[NP_{PL}[X\ N_{0,HEAD}[\text{INSTITUTION}(x)]\ Y] \rightarrow \text{STAFF\_MEMBERS}(x)]$

There is no evidence in (20)–(29) of probabilistic representation being required in the lexicon. The different interpretations are indicated through morphosyntactic choices.

Allan 1976; 2001 identify a principle of N_0 usage for English, given in (30).

(30) N_0, the form of the noun unmarked for number, is used when the denotation for N is perceived not to consist of a number of significant similar units.

In a plural NP headed by N_0, the absence of plural inflexion on the head noun marks 'collectivizing'. Collectivizing signals hunting, conservation, or farming jargon because N_0 is characteristically used of referents that are NOT perceived to be significant as individuals. Early users of the collectivized form were not interested in the individual animals except as a source for food or trophies. Consider the italicized nouns in (18)–(19) and (31)–(34), to which italics have been added.

(31) A three month shooting trip up the White Nile can offer a very good mixed bag, including, with luck, *Elephant*, *Buffalo*, *Lion*, and two animals not found elsewhere: Nile or Saddle-back (Mrs. Gray's) *Lechwe* and White-eared *Kob*. (Maydon (ed.) 1951: 168)

(32) On the way back to camp we sighted two *giraffe* on the other side of the river, which were coming down to the water's edge to drink. (Arkell-Hardwicke 1903: 285)

(33) These *cucumber* are doing well; it's a good year for them.

(34) The *cat-fishes*, of which there are about fifty distinct forms arranged in four families, constitute the largest group, with probably the greatest number of individuals per species. In some parts of the country where nets are little used and fishing is mainly done with traps and long lines, at least three-quarters of the annual catch is of *cat-fish*. (Welman 1948: 8)

The plural NP "cat-fishes" at the beginning of (34) refers to species of cat-fish whereas the N_0 at the end refers to individuals caught by fishermen. Collectivizing of trees and other plants is much less common than collectivizing animals – from which, perhaps, it derives. Vermin are never collectivized; though individual language users may differ over what counts as vermin. Early uses of the collectivized form were applied to animals hunted for food or trophies. Today, collectivizing occurs in contexts and jargons of hunting, zoology, ornithology, conservation, and cultivation where N_0 is characteristically used of referents that, as I've already said, are not perceived to be significant as individuals. Two possible contributing factors to the establishment of N_0 as the mark of collectivizing are (1) the unmarked plural of *deer* – which once meant "wild animal, beast", and (2) the fact that meat nouns are N_0 (discussed in the next section). Despite the fact that there is a good deal of variation in the data (see Allan 1976: 100f), collectivizable nouns should be marked as such in the lexicon. Reference will need to be made to the discourse domain being one of the contexts identified above and vermin will need to be excluded. The kind of entry I envisage is (35), which uses *giraffe* as an example.

(35) IF Domain = conservation THEN $\forall$ x[NP_{PL}[X N_0[GIRAFFE(x)] Y]]; CRED $\approx$ 0.6

Clearly, more work is needed.

6. Animals for food and fur

In this section I take up a discussion from Allan 1981. Consider the sentences in (36)–(37).

(36) Harry prefers lamb to goat.

(37) Jacqueline prefers leopard to fox.

Most likely you will interpret the animal product nouns in (36) to refer to meat, such that (36) is paraphrasable by (38), whereas the animal product nouns in (37) refer to animal pelts such that (37) is paraphrasable by (39).

(38) Harry prefers eating lamb to eating goat.

(39) Jacqueline prefers leopard skin to fox fur.

The converse interpretations are unlikely, especially *Jacqueline prefers eating leopard to eating fox*.[11] The predicate *prefer* in (36)–(37) offers a neutral context permitting the default animal product to rise to salience. This suggests that the lexicon entries for *lamb* and *goat*, and that for other creatures (such as *whale*, see (40)) should include a specific application of the formula in (41).

(40) In Tokyo, whale gets ever more expensive!

(41) $$\forall x \left[\begin{array}{l} \lambda y[NP_{MASS}[N(y) \wedge \text{ANIMAL}(y)](x)] \rightarrow \text{PRODUCT\_OF}(x) \\ \lambda y[NP_{MASS}[N(y) \wedge \text{ANIMAL}(y)](x)] \; +\!> \; \text{MEAT\_OF}(x) \end{array} \right]$$

The lexicon entries for *leopard* and *fox* should include a specific application of the formula in (43); so will all of the italicized animal product nouns in (42).

(42) (a) Jacqueline was wearing *mink*.

(b) Elspeth's new handbag is *crocodile*, I think.

(c) This settee's made of *buffalo*.

(d) The tannery has loads of *impala* right now.

(43) $$\forall x \left[\begin{array}{l} \lambda y[NP_{MASS}[N(y) \wedge \text{ANIMAL}(y)](x)] \rightarrow \text{PRODUCT\_OF}(x) \\ \lambda y[NP_{MASS}[N(y) \wedge \text{ANIMAL}(y)](x)] \; +\!> \; \text{PELT\_OF}(x) \end{array} \right]$$

11. I could find no online or corpora references to leopard meat or fox meat, but an Illinois butcher does offer lion meat, http://www.czimers.com/2.html (accessed July 14, 2010).

A mass NP headed by an animal noun will refer to the pelt of the animal denoted by that NP when there is in the clause an NP head or clause predicate describing apparel, accessories to apparel, furniture, the creation of an artefact, or any object likely to be made from leather and any place or process that involves pelts, hides, or leather such that these constrain the domain for the interpretation of N_0. Thus the nonmonotonic inference in (41) is cancelled by the implications of *the lining* in (44); from (43) the NMI is cancelled by the predicate *eat* in (45).

(44) I prefer the lining to be made of lamb, because it's softer.

(45) All we had to eat was leopard.

More subtle interpretations are required in (46)–(49).

(46) A plate of lamb can be worn by no-one.

(47) The girl holding the plate was wearing rabbit.

(48) The girl who wore mink was eating rabbit.

(49) Because she decided she preferred the lamb, Hetty put back the pigskin coat.

In (46) "plate of lamb" identifies meat. Although the most likely interpretation of *a plate of steel* is "a plate made of steel" (CRED $\geq$ 0.99), *a plate of lamb* is, with similar credibility, interpreted as "a plate bearing food". The predicate "wearing rabbit" in (47) identifies the rabbit pelts as apparel (again, CRED $\geq$ 0.99) and, likewise, "wore mink" in (48) identifies mink as apparel while the predicate in "eating rabbit" coerces the reference to rabbit meat. In (49) "the lamb" is most likely to be interpreted as meat (CRED $\geq$ 0.8) until this is revealed as a 'garden-path' misinterpretation corrected by the preference for a porcine pelt in the second clause which cancels this NMI, replacing it with the coerced interpretation "lambskin coat".

In this section I have claimed that animal nouns in mass NPs which denote a product from the dead animal typically refer to either the animal's flesh or its pelt, but this probabilistic inference can be cancelled by certain contextual elements that condition the domain for interpretation. Credibility rankings can be assigned as shown in (50). However, in (50) these rankings are based on my intuition, although they ought to be made on the basis of the frequency of interpretations retrieved from large and diverse corpora.

(50) NP_{MASS} [N[λy[LAMB(y) ∧ ANIMAL(y)](x)]] +> MEAT_OF(x); CRED ≥ 0.8
IF NOT MEAT_OF(x) THEN PELT_OF(x)

NP_{MASS} [N[λy[GOAT(y) ∧ ANIMAL(y)](x)]] +> MEAT_OF(x); CRED ≥ 0.7
IF NOT MEAT_OF(x) THEN PELT_OF(x)

NP_{MASS} [N[λy[RABBIT(y) ∧ ANIMAL(y)](x)]] +> MEAT_OF(x); CRED ≥ 0.7
IF NOT MEAT_OF(x) THEN PELT_OF(x)

NP_{MASS} [N[λy[LEOPARD(y) ∧ ANIMAL(y)](x)]] +> PELT_OF(x); CRED ≥ 0.9
IF NOT PELT_OF(x) THEN MEAT_OF(x)

NP_{MASS} [N[λy[FOX(y) ∧ ANIMAL(y)](x)]] +> PELT_OF(x); CRED ≥ 0.9
IF NOT PELT_OF(x) THEN MEAT_OF(x)

NP_{MASS} [N[λy[MINK(y) ∧ ANIMAL(y)](x)]] +> PELT_OF(x); CRED ≥ 0.9
IF NOT PELT_OF(x) THEN MEAT_OF(x)

NP_{MASS} [N[λy[BUFFALO(y) ∧ ANIMAL(y)](x)]] +> PELT_OF(x); CRED ≥ 0.8
IF NOT PELT_OF(x) THEN MEAT_OF(x)

NP_{MASS} [N[λy[CROCODILE(y) ∧ ANIMAL(y)](x)]] +> PELT_OF(x); CRED ≥ 0.8
IF NOT PELT_OF(x) THEN MEAT_OF(x)

NP_{MASS} [N[λy[IMPALA(y) ∧ ANIMAL(y)](x)]] +> PELT_OF(x); CRED ≥ 0.7
IF NOT PELT_OF(x) THEN MEAT_OF(x)

It would seem obvious that there should be some generalization over nouns that can refer to either meat or pelts; one might refer to the degree of choice between these two alternatives being "graded salience" (Giora 2003: 10 and this volume), but this notion is yet more relevant in the lexicon entry for *and*.

7. And

And may conjoin all sorts of sentence constituents and whatever is felicitously conjoined is grouped together such that there is always some plausible reason for the grouping. This 'plausibility' valuation is a coherence metric and necessarily pragmatic because it relies on knowledge of whatever world is spoken of; later, I shall question whether it is relevant to the

lexicon entry for *and*. With the exception of some conjoined NPs that I will refer to as NP-*COM-Conjunction (and briefly exemplify in (61)–(65)), the conjoined constituents are synonymous with a conjunction of sentences, e.g. in (51)(e) 'Two is a number ∧ Three is a number'.

(51) (a) Sue is tall and slim.

(b) Eric was driving too fast and hit a tree.

(c) Elspeth always drove slowly and carefully.

(d) Joe and Harriet are tall.

(e) Two and three are numbers.

On the assumption that Φ and Ψ are well-formed (combinations of) propositions expressed as well-formed conjunctions in English, the semantics of Φ *and* Ψ is as presented in (52). There is, in addition, a series of nonmonotonic inferences that exemplify Giora's "graded salience" (Giora 2003: 10); they are listed with the strongest contextually possible inference as the first to be considered.

(52) Φ *and* Ψ ↔ Φ ∧ Ψ

(a) IF CRED(¬Φ → ¬Ψ) ≥ 0.9 ∧ CRED(CAUSE(Φ,Ψ) ≥ 0.8

THEN Φ *and* Ψ +> Φ *causes* Ψ (e.g. *Flick the switch* ***and*** *the light comes on*; cause < effect[12]) ELSE

(b) IF CRED(ENABLE ([DO(Ø,Φ)],Ψ)) ≥ 0.9 ∧ CRED(¬Φ → ¬Ψ) ≥ 0.8

THEN Φ *and* Ψ +> Φ *enables the consequence* Ψ ∨ Φ *is a reason for* Ψ (e.g. *Stop crying* ***and*** *I'll buy you an ice-cream*; action < consequence) ELSE

(c) IF CRED(ΦΨ) ≥ 0.8

THEN Φ *and* Ψ +> Φ *and then later* Ψ (e.g. *Sue got pregnant* ***and*** *married her boyfriend*; Φ < Ψ) ELSE

12. Φ < Ψ means "Φ precedes Ψ (chronologically)"

(d) IF CRED(ENABLE(Φ,[DO(S,[SAY(S,Ψ)])])) ≥ 0.8[13]

THEN Φ *and* Ψ +> Φ *is background for* Ψ (e.g. *There was once a young prince,* ***and*** *he was very ugly*) ELSE

(e) Φ *and* Ψ +> Φ *is probably more topical or more familiar to S than* Ψ (e.g. *On Saturdays my mum cleans the flat* ***and*** *Sue washes the clothes*)

Note the conditional relations in (53):

(53) (Φ *causes* Ψ) → (Φ *is a reason for or enables the consequence* Ψ) → (Φ *temporally precedes* Ψ)[14]

Whether the last two discourse based implicatures of (52) are part of this sequence remains to be determined. However, it is arguable that if Φ *is background for* Ψ then Φ is prior to Ψ; and if Φ *is more topical or more familiar than* Ψ, then again, it is arguable that Φ is prior to Ψ; and should these rather tenuous claims be acceptable, then the fact that Φ precedes Ψ when they are conjoined is normally iconic. However, the choice of sequence is a matter of usage (or pragmatics) and is not obligatory, but it does seem to justify a general statement such as (54):

(54) Φ *and* Ψ ↔ Φ ∧ Ψ

Φ *and* Ψ +> Φ is prior to Ψ; CRED ≥ 0.9

Consider (from (52)) *Sue got pregnant and married her boyfriend*: it is false (CRED = 0) that Sue's getting pregnant literally causes her to marry her boyfriend, though it may be her reason for doing so, CRED ≈ 0.4; but it is quite probable (CRED ≈ 0.75) that her marriage to the boyfriend is a consequence of her being pregnant, whether or not he is the biological father-to-be. It is almost certain (CRED ≥ 0.9), even though defeasible, that Sue's pregnancy precedes her marriage. Out of any natural context of use it is not possible to determine whether or not saying *Sue got pregnant* is a background for going on to say that she married her boyfriend.

13. S identifies the speaker, here and below.

14. Kasia Jaszczolt (p.c.) has questioned whether temporal precedence is applicable with statives such as *She is underage and can't drive*. I don't strongly disagree but I think being underage is prior to inability to drive and this is evident in *She is no longer underage and can now drive.*

This aside, it has been possible to propose a (partial) lexicon entry for *and* which includes its implicatures in grades of salience. There seems to be no good reason to treat *and* as multiply ambiguous semantically when one core meaning can be identified (logical conjunction) and all other interpretations can be directly related to that as a hierarchy of nonmonotonic inferences processed algorithmically. As Ockham wrote: *Numquam ponenda est pluralitas sine necessitate* 'Plurality should never be posited without necessity' (Ordinatio Distinctio 27, Quaestio 2, Ockham 1967-88: I, K)

Is it possible to define a plausibility measure for Φ *and* Ψ that is semantically based? I suspect not. At first sight the acceptability of (55) as against the unacceptability of (56) seems explicable semantically because only living things eat and if Max is dead he is no longer living and this is semantic entailment of *die*.

(55) Max ate a hearty meal and died.

(56) *Max died and ate a hearty meal.

However, the situation seems pragmatically determined in (57)–(60): it is a matter of conventional beliefs about death, going to hospital, and going to heaven.

(57) Max went to hospital and died there.

(58) *Max died and went to hospital.

(59) Max died and went to heaven.

(60) *Max went to heaven and died there.

In NP-*COM-Conjunction, *COM is a ≥2-place predicate with a sense "is added to, is mixed or combined with, acts jointly or together with, is acted upon jointly or together with" (Allan 2000: 196). It is found in (61), which is not semantically equivalent to (62) – contrast the latter with (51)(e).

(61) Two and three are five.

(62) *Two is five ∧ Three is five

A revealing recipe-like paraphrase of (61) is (63), which accounts for the fact that (64) is a paraphrase of (61).

(63) Take two$_x$ and take three$_y$, combine them (*COM(x,y)), and you get five$_w$, cf. *Mix flour$_x$ and water$_y$ to make paste$_w$* or just *Flour and water make paste*.

(64) Two and three make five.

NP-*COM-Conjunction is recognized when a conjunction of sentences either cannot apply or is unlikely to apply as in (61) and (65).

(65) Joe and his wife have a couple of kids.

The subject NP of (65) is most likely NP-*COM-Conjunction whereas that of (66) is not. That these judgments are pragmatically rather than semantically plausible is seen by comparing them.

(66) Joe and his sister have a couple of kids.

(66) is, given social constraints on incest, most likely an infelicitous manner of expression where the conjunction is intended to be Φ *and* Ψ with the weakest of nonmonotonic inferences; preferred would be *Joe and his sister each have a couple of kids*. With respect to (65), although it is true that each of Joe and his wife has two kids, the sentence *Joe and his wife each have a couple of kids* suggests these derive from former relationships such that the married couple has four children altogether.

8. Sorites

Two horses don't constitute a *herd* nor do ten grains of sand constitute a *heap*. For collections such as these, denoted by sorites[15] nouns, the number of constituents needed to render the description accurate depends on the nature of the constituents: for example, whereas the least lower bound on a herd of horses might be three, that on a heap of sand is probably more than a hundred. There are sorites predicates like *be bald, be tall, be many* and sorites adverbs like *slowly, loudly.* These are invariably gradable and contextually determined as may be seen from the contrasts in (67).

15. *Sorites* from Greek σωρείτης "heaped up". The earliest discussion of sorites paradoxes is attributed to Eubulides of Miletus, 4th century BCE. A single grain of sand is certainly not a heap. Nor is the addition of a single grain of sand enough to transform a non-heap into a heap. If we keep adding grains, at some point we will have a heap – but there is no agreement on the precise number that constitutes the least lower bound of a heap.

(67) ***tall*** *for a Pygmy* VERSUS ***tall*** *for a North American basket-ball professional*[16]

many *people thought George W Bush was a fool* VERSUS ***many*** *of my students didn't attend class today*

a slug moves ***slowly*** VERSUS *the train went through the station* ***slowly***

There is a similar contextual relevance for the nouns: a *herd* of horses, elephants or giraffe will typically have fewer members than a herd of wildebeest, though this is not necessarily the case; moreover, it has no bearing on the lexical meaning of *herd*. The least lower bound on a *heap* of beans is lower than that on a *heap* of sand, probably because of the size of the constituent members. Clearly these are facts about the world referred to but are they facts about the meaning of listemes? No, but they are relevant to the propositions in which the listemes occur: for instance, if speakers wish to report the speed at which a slug is moving they need to apply different criteria than when reporting the speed at which a train is moving. It appears from work reported by Hagoort, Hald, Bastiaansen et al. 2004 that the brain is prepared to do exactly that kind of thing and that contextual information is integrated with semantic information from the start, see also Terkourafi 2009. However, as I've said, although this is relevant to the meaning of propositions, we can dispense with such enriched interpretations in the lexicon because they are instances of lexical adjustment: they count as 'ad hoc categories' (Barsalou 1983; Carston 2002; Wilson and Carston 2007) dependent on a particular domain of discourse. What we see in (67) is a context induced specification of the meaning for the sorites words. The same holds for *bald*: various degrees of baldness are characterized in (68)–(70).

(68) His hair is thinning / thin ≈ He is balding / going bald / has a bald patch.

(69) He is bald.

(70) He is completely bald.

The domain of baldness extends from thinning (head) hair to its almost complete absence. It is arguable that (69) is applicable in situations where (68) or else (70) would also hold true,

16. The average height for a male pygmy is less than 5′ (155 cm, http://www.physorg.com/news117456722.html); for a basket-ball player it is 6′6″ (198 cm; http://wiki.answers.com/Q/What_is_the_average_height_of_a_basketball_player).

even though the accuracy of (69) might be disputed in favour of either (68) or (70). So, how sorites words should be specified in a lexicon is highly controversial.

Although not directly concerned with the lexicon, there is a large number of proposals discussed in Williamson 1994; Beall (ed.) 2003 and Smith 2008. They include supervaluation, subvaluation, and plurivaluation. Smith suggests "talk of the meanings of some terms must always be relative to a group of speakers, whose dispositions regarding the use of those terms plays an essential part in fixing those meanings" (Smith 2008: 314). This is a recasting of Quine's "There is nothing in linguistic meaning beyond what is to be gleaned from overt behavior in observable circumstances" (Quine 1992: 38). To return to (69): what I suggest for the meaning of *bald* is the minimal semantics of (71).

(71) $\text{BALD}(x) \rightarrow \neg[\text{FULL\_COMPLEMENT\_OF\_HAIR}(x)]$

Two speakers, or the same speaker on different occasions, may differ as to what counts as 'not a full complement of hair' such that *x is bald* has a range of truth values; i.e. there is no single state of hair-loss for which it is invariably true of x that *x is bald* for all occasions and all speakers. A modification like (68) is appropriate to the least lower bound and (70) to the greatest upper bound; (69) applies to both.

Defining sorites terms often invokes alternative points on the relevant scale. For instance *many* implies a contrast with other points on a quantity scale; more precisely, less than *most* and greater than *a few*. In (72), $\mid f \cap g \mid$ can be glossed 'the number of Fs that (are) G'.

(72) $[\text{MANY}(x)\text{: } Fx](Gx) \rightarrow \mid f \cap g \mid > [\text{A\_FEW}(x)\text{: } Fx]G(x)$

$+> \mid f \cap g \mid < [\text{MOST}(x)\text{: } Fx]G(x)$

(I assume that *a few x* > *few x* > *one x*.) The domain referred to significantly affects the actual numbers, as we saw in (67). It is notable that to establish the truth of (73) we cannot look to a specific number because even if that can ever be known, the precise number that justifies the use of "many" will differ for different speakers and even for the same speaker on different occasions.

(73) Many US citizens live in poverty.

Although the meaning of (73) falls under the definition in (72) there is also an implication, or perhaps connotation, that (according to the speaker) *the number of US citizens living in poverty is greater than it ideally ought to be.* Similar conditions hold for *Many of my students were absent from class today* which does not imply that more than half of them weren't there, but that 'more than one might have expected to be absent were in fact absent' – and that could easily be as little as 5%.

For sorites like *tall* and *slowly* it will be necessary to invoke, respectively, the height scale *tall* > *average height* > *short* and the speed scale *slow* < *average speed* < *fast* on condition that these apply to a particular domain or set of domains as shown in (67).

Sorites like *herd* and *heap* (in the sense of Eubulides' *soros*) involve configurational criteria.

(74) $$\forall x,y \left[\begin{array}{l} \text{HERD}(x) \text{ of } c \rightarrow c = \{y: y \text{ is a member of } x\} \\ \qquad\qquad \wedge \text{ TRAVEL\_TOGETHER}(y) \wedge \ |c| \geq 3 \\ \text{HERD}(x) \text{ of } c \ +> \ |c| \gg 3 \end{array}\right]$$

(75) $$\forall x,y \left[\begin{array}{l} \text{HEAP}(x) \text{ of } c \rightarrow c = \{y: y \text{ is a constituent of } x\} \\ \qquad\qquad \wedge \text{ COLLOCATED\_INTO\_A\_ROUGH\_CONE}(y) \wedge \ |c| \geq 3 \\ \text{HEAP}(x) \text{ of } c \ +> \ |c| \gg 3 \end{array}\right]$$

Suppose that three is the least lower bound for a herd or heap and often the number of constituents is many more, often vastly many more. There is no upper bound. A *heap of sand* will typically have many more constituents than a *heap of logs*; though if the domain of discourse is an egg-timer on the one hand and a clear-felled forest on the other, there may not be such a discrepancy. There is no unique quantity that defines a heap, not even a heap of some particular substance; that is, there is no exact number that determines when a quantity of sand constitutes a heap; the roughly-conical configuration is a necessary part of the requirement but is insufficient in itself – as is the condition on quantity. However, the semantic extension of *heap(s)* as in *I have a heap of things to do* and *There were heaps of people at the party* has lost all notion of a particular configuration and is roughly synonymous with *lots of* or *many* and must be defined in a manner similar to (72).

9. Formulaic language in the lexicon

"A formulaic sequence is a sequence, continuous or discontinuous, that appears prefabricated and stored as a chunk, rather than being generated afresh" (Wray 2008: 94). Just as metaphor is pervasive in language, so are "prefabs" – a useful term succinctly defined by Erman and Beatrice 2000 as "specific conventionalized multiword strings". Especially in the spoken language, people use thousands of them (just look, for example, at http://www.phrases.org.uk/index.html); but they are also markers of oral literature, religious texts, best-seller scripts, and popular radio and TV shows (see Allan 2001; 2006; Corrigan, Moravcsik, Ouali et al. (eds) 2009; Donahue 1991; Goldman 1990; Jackendoff 1995; Jensen 1980; Kuipers 2009; Paraskevaides 1984; Schmitt 2004; Wray 2002; 2008). Prefabs can be classed into at least three groups.

Idioms are primarily figurative; they include: *a bit of the other; Bob's your uncle; by and large; come a cropper; fuck off; go the whole hog; kick the bucket; put a sock in it; rain cats and dogs; set store by; sleep like a log; spill the beans; sweat blood; the key to*.

Clichés are primarily nonfigurative; they include: *be heavily compromised; be not very well; believe you me; don't do anything I wouldn't do; Good Lord; Happy Birthday! Hot-dog!* [= great!]*; ladies and gentlemen; out of sight out of mind; reading, writing, and* (*a*)*rithmetic; to make a long story short; un je ne sais quoi; you can say that again; you'd better* [do A].

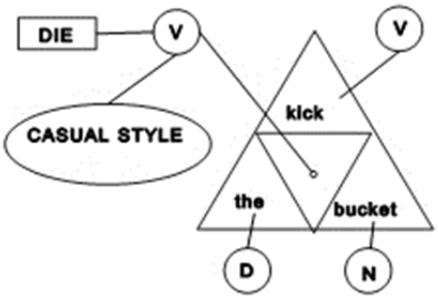
Catch-phrases include: *Beam me up, Scotty; Computer says 'No'; Frankly, my dear, I don't give a damn; It doesn't amount to a hill of beans; Not that there's anything wrong with it; One potato, two potato, three potato, four …; Play it again Sam; S/he loves me, s/he loves me not*.

Subclassifications of these groups sometimes suggest themselves (e.g. imprecations, proverbs) and a prefab can often be classed into more than one of the three (e.g. *be worth one's weight in gold*).

Prefabs have similar characteristics to compounds and phrasal verbs in that, although they may have a variable slot, they are largely immutable and function as lexical islands phonologically and syntactically (Van Lancker, Canter and Terbeek 1981; Underwood, Schmitt and Galpin 2004; Wray 2002; 2008). Like proper names and tabooed terms (such as *fuck*) they seem to be stored in a different manner from the normal lexicon, perhaps in the right brain. The evidence for this is that people with left hemisphere trauma often have access to prefabs, proper names, and tabooed terms when they don't have normal access to ordinary language; furthermore, persons with right hemisphere damage use significantly fewer prefabs than normal subjects (Van Lancker Sidtis 2009: 452). Lexicography has ignored the conclusion that different kinds of vocabulary are stored in different hemispheres of the brain, even though it could be relevant to classifying types of lexical data; I shall maintain this tradition.

A simplified lexicon entry for *kick the bucket* might be something like (76).

(76) /kɪk ðə bʌkət/ — [VP[V[KICK]] NP[D[THE] N[BUCKET]]] → DIE(x)



The ellipse in the figure contains encyclopaedic information that is clearly pragmatic yet according to Allan 2001 is outside of the lexicon. Traditionally such information is located in dictionaries, for instance, the *Oxford English Dictionary* 1989 labels *kick the bucket* "Slang" and the *Macquarie Dictionary* 2003 describes it as "Colloquial" (it doesn't appear in Webster 2002). Such descriptions, whether assigned to the lexicon or the networked encyclopaedia, are clearly pragmatic. The explanation for the meaning of *kick the bucket* is metonymic: in former times a *bucket* was a 'beam' and when an animal (such as a pig) was tied to the beam by its hind legs to be slaughtered, it would kick the bucket in the throes of

death. But information about this source for the idiom is an encyclopaedic datum that is not generally known, and plays no part in the interpretation today of the idiom *kick the bucket*.

Unlike the meaning of the typical idiom, the meaning of a typical cliché is computable from its constituent parts. What marks the cliché is that it occurs frequently as the clichéd chunk (Bannard and Lieven 2009: 300f, 304), and experimental evidence suggests that it is normally processed as a chunk and not according to its constituent parts (Underwood, Schmitt and Galpin 2004, Wray 2002; 2008). I suggest that clichés should therefore be noted in full in a lexicon and (pragmatically) marked as clichés. Mutatis mutandis, the same goes for catch-phrases: their meaning is almost invariably computable from their parts, but they are recalled and used as chunks – or perhaps as articulated chunks in the case of items of play like *one potato, two potato, three potato, four...*, or the words of a national anthem or of the full version of *Happy Birthday to you ....* It is a debatable matter whether these can count as lexical entries rather than encyclopaedia entries. They seem to be evoked by a particular kind of event that triggers a speech act, e.g. *happy birthday* by the occasion of someone's birthday that the speaker wishes to demonstrably recognize; *Beam me up, Scotty* is triggered by the thought 'Get me out of here'. It seems feasible to propose that the listeme *birthday* is linked to the networked encyclopaedia with a free pragmatic condition like (77):

(77) If it is X's birthday then it is appropriate to tell X *Happy birthday.*

The situation with respect to *Beam me up, Scotty* is far more constrained: it can perhaps be tagged to the phrasal verb *get NP out* in some thesaurus-like way on condition that the constituent NP refers to the speaker (perhaps, along with others); it can only be used as a jocular expression and to an addressee likely to understand the utterance as a catch-phrase. This latter condition does not apply to all catch-phrases: for instance, it doesn't apply to *not that there is anything wrong with it* which functions adequately as a non-prefab; the condition that applies is that "it" refers to a mildly tabooed topic (such as being gay)[17]. This illustrates the squishiness[18] of prefabs.

17. "Seinfeld" Season 4, Episode 17 "The Outing" (1993).
18. After Ross 1972.

Prefabs are, by definition, multiword expressions. Traditional dictionaries of phrases list them in alphabetical order but the mental lexicon is surely more akin to a database which is searched in a manner similar to a Google search engine operating on key words and combinations of words. The mental lexicon will also be accessed semantically and pragmatically (i.e. via meanings and encyclopaedic information, see Giora this volume and Katsos this volume) and not merely through aspects of the form of language expressions.

10. Connotation in the lexicon

The connotations of a language expression are pragmatic effects that arise from encyclopaedic knowledge about its denotation (or reference) and also from experiences, beliefs, and prejudices about the contexts in which the expression is typically used. Terms like *surgeon, nurse, secretary/receptionist* and *motor mechanic* evoke connotations of gender from the fact that the typical job-holder in each case is, even today, a gendered stereotype: most surgeons and motor mechanics are male; most nurses and secretary/receptionists are female. These connotations are all, clearly, the pragmatic effects of normative conceptions of typical job-holders.

(78) *surgeon* → a medical practitioner who treats wounds, fractures, deformities, or disorders by manual operation and/or instrumental appliances
+> a male medical practitioner who treats wounds, fractures, deformities, or disorders by manual operation and/or instrumental appliances, CRED ≈ 0.85

(79) *nurse* → a person employed or trained to take charge of a young children or who cares for the sick or infirm
+> a woman employed or trained to take charge of a young children or who cares for the sick or infirm, CRED ≈ 0.94

The most common denotations of *bunny* and *rabbit* or *doggie* and *dog* are the same, but the connotations are different: bearing the diminutive, the first member of the two pairs connotes endearment or childish language; see (80).

(80) *doggie* → dog

+> the speaker is a child ∨ the speaker is addressing a child with respect to the animal ∨ the speaker is expressing endearment with respect to the animal

To avoid blaspheming (for which the Bible sanctions execution, Leviticus 24: 16), people use a variety of euphemistic expletives (see Allan and Burridge 2006: 15ff, 39). For instance, *Jesus* is end-clipped to *Jeeze!* and *Gee!* (which is also the initial of *God*); *Gee whiz!* is a remodelling of either *jeeze* or *jesus*. More adventurous remodellings are *By jingo! Jiminy cricket!* [from *Jesus Christ*] *Christmas! Crust! Crumbs! Crikey!* Note that the denotation of *Gee! Jeepers!* and *Jesus!* is identical. All function as exclamations of surprise, dismay, enthusiasm, or emphasis. From a purely rational viewpoint, if one of them is blasphemous, then all of them are. What is different is that the first two have connotations that are markedly different from the last. Connotation – or, more precisely its pragmatic effect, reaction to connotation – is seen to be a vocabulary generator. But the question here is what goes into the lexicon, and I suggest (81)–(82) (in which statements introduced by a simple + are encyclopaedic).

(81) *Jesus* → Proper name for a male

+> Jesus Christ of Nazareth, son of Mary (Mariam)

\+ Jesus the Christ or Messiah, central figure of Christianity which takes him to be the son of God.

Jesus → Interjection (expressive idiom). Blasphemous exclamation of surprise, dismay, enthusiasm, or emphasis.

+> Often not regarded as literally blasphemous, CRED ≈ 0.8

(82) *Jeepers* → Interjection (expressive idiom). Exclamation of surprise, dismay, enthusiasm, or emphasis.

\+ Euphemism based on remodelling of the blasphemy *Jesus*.

Whether the encyclopaedic statements should be included within the lexicon is a matter of debate. I personally don't believe they should form a part of the lexicon entry but they must certainly be accessible from and networked with the lexicon.

11. Conclusion

In this chapter I have looked at ways in which pragmatics intrudes on the lexicon. I count as "pragmatic" encyclopaedic data and nonmonotonic inferences (NMI) – which arguably arise from encyclopaedic data. In §2, I introduced the notion of a credibility metric for a proposition and used it to calibrate NMIs in the lexicon to correspond with the degree of confidence one might have in the truth of the inference: its probability. §3 and §4 demonstrated that in addition to the lexicon entry specifying the necessary components of meaning in the semantics for an entry, it should also specify the most probable additional components of meaning, which are accepted or cancelled as a function of contextual constraints. These same sets of conditions were demonstrated for different kinds of entries throughout the rest of the chapter. §5 looked at lexicon entries for collective and collectivizable nouns. These differ in that different interpretations for collective nouns arise from their morphosyntactic context and although this needs to be captured in the lexicon it is not a matter of pragmatics; on the other hand, a noun is collectivizable only in some defined set of contexts and these are a pragmatic constraint. §6 discussed the use of animal nouns in mass NPs to denote either the animal's meat or its pelt. Although there are defined morphosyntactic conditions on such interpretations, the choice of one interpretation or the other is pragmatically determined because it is contextually induced and is open to calibration against a credibility metric. §7 returned to the much disputed semantics of *and*. The view taken here is for a monosemic semantics which assumes English *and* has the semantics of logical conjunction but there is a graded salience captured in an algorithm that assigns one of a set of nonmonotonic inferences as supplementary meaning on the basis of context. §8 discussed the vexed question of how to represent the semantics of sorites terms in the lexicon. A minimalist semantics was proposed. §9 discussed the matter of prefabs or formulaic expressions. It is only recently that their frequency and ubiquity has been recognized. They

pose a challenge to the lexicon principally because they are multiword expressions; many are figurative; many are stylistically marked. These pragmatic characteristics are appropriate to encyclopaedic information linked to the entry. §10 considered the representation of connotation in the lexicon as a matter of pragmatic intrusion.

In this chapter I have shown different motivations for including pragmatics in the lexicon or linking it to the lexicon, and I have demonstrated how that may be accomplished. This is not to deny that other formalizations are possible.

References

Allan, Keith. 1976. Collectivizing. *Archivum Linguisticum* 7: 99-117.

Allan, Keith. 1980. Nouns and countability. *Language* 56: 541–67. (http://www.jstor.org/stable/414449).

Allan, Keith. 1981. Interpreting from context. *Lingua* 53: 151–73. (http://dx.doi.org/10.1016/0024-3841(81)90015-2).

Allan, Keith. 2000. Quantity implicatures and the lexicon. In *The Lexicon-Encyclopedia Interface*, ed. by Bert Peeters. Amsterdam: Elsevier. Pp. 169–218.

Allan, Keith. 2001. *Natural Language Semantics*. Oxford & Malden MA: Blackwell.

Allan, Keith. 2006. Lexicon: structure. In *Encyclopedia of Languages and Linguistics. 2nd edn*, ed. by E. Keith Brown. 14 vols. Oxford: Elsevier. Pp. 7: 148-51.

Allan, Keith and Kate Burridge. 2006. *Forbidden Words: Taboo and the Censoring of Language*. Cambridge: Cambridge University Press.

Arkell-Hardwicke, A. 1903. *An Ivory Trader in North Kenya*. London: Longmans, Green and Co.

Bannard, Colin and Elena Lieven. 2009. Repetition and reuse in child language learning. In *Formulaic Language: Volume 2 Acquisition, Loss, Psychological Reality, and Functional Explanations*, ed. by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali and Kathleen M. Wheatley. Amsterdam/Philadelphia: John Benjamins. Pp. 299-321.

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11: 211-27.

Bauer, Laurie. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.

Beall, J.C. (ed.) 2003. *Liars and Heaps: New Essays on Paradox*. Oxford: Clarendon Press.

Berlin, Brent and Paul Kay. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley and Los Angeles: University of California Press.

Blutner, Reinhard. 1998. Lexical pragmatics. *Journal of Semantics* 15: 115-62.

Blutner, Reinhard. 2004. Pragmatics and the lexicon. In *The Handbook of Pragmatics*, ed. by Laurence R. Horn and Gregory Ward. Malden MA: Blackwell. Pp. 488-514.

Blutner, Reinhard. 2009. Lexical pragmatics. In *The Pragmatics Encyclopedia*, ed. by Louise Cummings. London: Routledge.

Carston, Robyn. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford & Malden MA: Blackwell.

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 (ACL97), Madrid, July 7th-12th 1997.* (Madrid) Pp. 136-43.

Corrigan, Roberta L., Edith A. Moravcsik, Hamid Ouali and Kathleen M. Wheatley. (eds) 2009. *Formulaic Language*. 2 vols. Philadelphia: John Benjamins.

Di Sciullo, Anna-Maria and Edwin Williams. 1987. *On the Definition of Word*. Cambridge MA: MIT Press.

Donahue, Dennis P. 1991. *Lawman's Brut, An Early Arthurian Poem: A Study of Middle English Formulaic Composition*. Lewiston NY: Mellen.

Erman, Britt and Warren Beatrice. 2000. The idiom principle and the open choice principle. *Text* 20: 29-62.

Fillmore, Charles J. 1982. Frame semantics. In *Linguistics in the Morning Calm*, ed. by Linguistic Society of Korea. Seoul: Hanshin. Pp. 111–38.

Fillmore, Charles J. 2006. Frame semantics. In *Encyclopedia of Languages and Linguistics. 2nd edn*, ed. by E. Keith Brown. 14 vols. Oxford: Elsevier. Pp. 4: 613-20.

Fillmore, Charles J. and Beryl T. Atkins. 1992. Toward a frame-based lexicon: the semantics of RISK and its neighbors. In *Frames, Fields, and Contrasts*, ed. by Adrienne Lehrer and Eva F. Kittay. Hillsdale: Lawrence Erlbaum. Pp. 75–102.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50. Reprinted as 'On sense and reference'. In *Translations from the Philosophical Writings of Gottlob Frege*, ed. by Peter Geach and Max Black. Oxford: Blackwell. 1960: 56–78.

Giora, Rachel. 2003. *On Our Mind: Salience, Context, and Figurative Language*. New York: Oxford University Press.

Goldman, Kenneth A. 1990. *Formulaic Analysis of Serbo-Croation Oral Epic Songs: Songs of Avdo Avdic*. New York: Garland.

Hagoort, Peter, Lea Hald, Marcel Bastiaansen and Karl M. Petersson. 2004. Integration of word meaning and world knowledge in language comprehension. *Science* 304: 438-41.

Haiman, John. 1980. Dictionaries and encyclopedias. *Lingua* 50: 329-57.

Hanks, Patrick. (ed.) 1979. *Collins Dictionary of the English Languagee*. London: Collins.

Huang, Yan. 2009. Neo-Gricean pragmatics and the lexicon. *International Review of Pragmatics* 1: 118-53.

Jackendoff, Ray S. 1985. Multiple subcategorization and the θ-criterion: the case of *climb*. *Natural Language and Linguistic Theory* 3: 271–95.

Jackendoff, Ray S. 1995. The boundaries of the lexicon. In *Idioms: Structural and Psychological Perspectives*, ed. by Martin Everaert, Erik-Jan van der Linden, André Schenk and Ron Schreuder. Hillsdale NJ: Erlbaum. Pp. 133–65.

Jensen, Minna S. 1980. *The Homeric Question and the Oral-Formulaic Theory*. Copenhagen: Museum Tusculanum Press.

Kernfeld, Barry. 1994. *The New Grove Dictionary of Jazz*. London: Macmillan.

Kuipers, Koenraad. 2009. *Formulaic Genres*. Basingstoke: Palgrave Macmillan.

Labov, William. 1978. Denotational structure. In *Papers from the Parasession on the Lexicon*, ed. by Donka Farkas, Wesley M. Jacobsen and Karol W. Todrys. Chicago: Chicago Linguistics Society. Pp. 220–60.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 75: 522-51.

MacLaury, Robert E. 1997. *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. Austin: University of Texas Press.

Macquarie Dictionary. 2003. 3rd edn, revised. North Ryde NSW: Macquarie Library.

Maydon, Hubert C. (ed.) 1951. *Big Game Shooting in Africa*. London: Seeley Service and Co.

McCawley, James D. 1975. Lexicography and the count-mass distinction. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*. Berkeley Berkeley Linguistics Society. Pp. 314–21. Reprinted in James D. McCawley *Grammar and Meaning*. Tokyo: Taikushan. 1973: 165-73.

McCawley, James D. 1978. Conversational implicature and the lexicon. In *Syntax and Semantics 9: Pragmatics*, ed. by Peter Cole. New York: Academic Press. Pp. 245-59.

Nunberg, Geoffrey and Annie Zaenen. 1992. Systematic polysemy in lexicology and lexicography. In *EURALEX '92: Proceedings I-II: Papers submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland*, ed. by Hannu Tommola, Krista Varantola, Tarja Salmi-Tolonen and Jürgen Schopp. Tampere: Tampereen yliopisto. Pp. 387-98.

Ockham, William. 1967-88. *Guillelmi de Ockham Opera Philosophica et Theologica* Ed. by Gedeon Gál and Stephen Brown. 17 vols. St Bonaventure NY: The Franciscan Institute St Bonaventure University.

Oxford English Dictionary. 1989. 2nd edn. Oxford: Clarendon Press. [Abbreviated to *OED*]. Also available on Compact Disc.

Paraskevaides, H.A. 1984. *The Use of Synonyms in Homeric Formulaic Diction*. Amsterdam: A.M. Hakkert.

Pearsall, Judy. (ed.) 1998. *New Oxford Dictionary of English*. Oxford: Oxford University Press.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge MA: MIT Press.

Quine, Willard V.O. 1992. *Pursuit of Truth*. 2nd revised edn. Cambridge MA: Harvard University Press.

Ross, John R. 1972. The category squish: Endstation Hauptwort. In *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*, ed. by Paul M. Peranteau, Judith N. Levi and Gloria C. Phares. Chicago: Chicago Linguistic Society. Pp. 316-38.

Schmitt, Norbert. 2004. *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam/Philadelphia: John Benjamins.

Smith, Nicholas J.J. 2008. *Vagueness and Degrees of Truth*. Oxford: Oxford University Press.

Stubbs, Michael. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Terkourafi, Marina. 2009. On de-limiting context. In *Contexts and Constructions*, ed. by Alexander Bergs and Gabriele Diewald. Amsterdam/Philadelphia: John Benjamins. Pp. 17-42.

Uhlig, Gustav. (ed.) 1883. *Grammatici Graeci*. Leipzig: Teubner.

Underwood, Geoffrey, Norbert Schmitt and Adam Galpin. 2004. The eyes have it: an eye-movement study into the processing of formulaic sequences. In *Formulaic Sequences: Acquisitin, Processing and Use*, ed. by Norbert Schmitt. Amsterdam/Philadelphia: John Benjamins. Pp. 153-72.

Van Lancker, Diana, Gerald J. Canter and Dale Terbeek. 1981. Disambiguation of ditropic sentences: acoustic and phonetic cues. *Journal of Speech and Hearing Research* 24: 330-35.

Van Lancker Sidtis, Diana. 2009. Formulaic and novel language in a 'dual process' model of language competence. In *Formulaic Language: Volume 2 Acquisition, Loss, Psychological Reality, and Functional Explanations*, ed. by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali and Kathleen M. Wheatley. Amsterdam/Philadelphia: John Benjamins. Pp. 445-70.

Vigliocco, Gabriela, Lotte Meteyard, Mark Andrews and Stavroula Kousta. 2009. Toward a theory of semantic representation. *Language and Cognition* 1: 219–47.

Webster, Noah. 2002. *Webster's Third New International Dictionary of the English Language, Unabridged*. Ed. by Philip B. Gove. Springfield MA: Merriam-Webster.

Weinreich, Uriel. 1966. Explorations in semantic theory. In *Current Trends in Linguistics 3*, ed. by Thomas A. Sebeok. The Hague: Mouton. Reprinted in *Weinreich on Semantics* ed. by William Labov and Beatrice S. Weinreich. Philadelphia: University of Pennsylvania Press. 1980: 99–201.

Welman, J.B. 1948. *Preliminary Survey of the Freshwater Fishes of Nigeria*. Lagos: Government Printer.

Williamson, Timothy. 1994. *Vagueness*. London: Routledge.

Wilson, Deirdre and Robyn Carston. 2007. A unitary approach to lexical pragmatics: relevance, inference and ad hoc concepts. In *Pragmatics*, ed. by Neil Burton-Roberts. Houndmills: Palgrave Macmillan. Pp. 230-59.

Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, Alison. 2008. *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge MA: Addison-Wesley.