The Algebra of Lexical Semantics

Harvard University,
1737 Cambridge St, Cambridge MA 02138
Hungarian Academy of Sciences,
Computer and Automation Research Institute,
13-17 Kende u, H-1111 Budapest
andras@kornai.com
http://kornai.com

Abstract. The current generative theory of the lexicon relies primarily on tools from formal language theory and mathematical logic. Here we describe how a different formal apparatus, taken from algebra and automata theory, resolves many of the known problems with the generative lexicon. We develop a finite state theory of word meaning based on *machines* in the sense of Eilenberg [11], a formalism capable of describing discrepancies between syntactic type (lexical category) and semantic type (number of arguments). This mechanism is compared both to the standard linguistic approaches and to the formalisms developed in AI/KR.

1 Problem Statement

In developing a formal theory of lexicography our starting point will be the informal practice of lexicography, rather than the more immediately related formal theories of Artificial Intelligence (AI) and Knowledge Representation (KR). Lexicography is a relatively mature field, with centuries of work experience and thousands of eminently usable work products in the form of both mono- and multilingual dictionaries. In contrast to this, KR is a rather immature field, with only a few decades of work experience, and few, if any, usable products. In fact, our work continues the trend toward more formalized lexicon-building that started around the Longman Dictionary (Boguraev and Briscoe [6]) and the Collins-COBUILD dictionary (Fillmore and Atkins [14]), but takes it further in that our focus is with the mathematical foundations rather than the domain-specific algorithms.

An entry in a standard monolingual dictionary will have several components, such as the etymology of the word in question; part of speech/grammatical category information; pronunciation guidelines in the form of phonetic/phonological transcription; paradigmatic forms, especially if irregular; stylistic guidance and examples; a definition, or several, for different senses of the word; and perhaps even a picture, particularly for plants, animals, and artifacts. It is evident from

C. Ebert, G. Jäger, and J. Michaelis (Eds.): MOL 10/11, LNAI 6149, pp. 174–199, 2010.

the typeset page that the bulk of the information is in the definitions, and this is easily verified by estimating the number of bits required to encode the various components. Also, definitions are the only truly obligatory component, because a definition will be needed even for words lacking in exceptional forms (these are the majority) or an interesting etymology, with a neutral stylistic value, predictable part of speech (most words are nouns), and an orthography sufficiently indicative of pronunciation.

There is little doubt that definitions are central to the description of words, yet we have far richer and better formalized theories of etymology, grammatical category, morphological structure, and phonological transcription than we have theories of word meaning. Of necessity, work such as Dowty [8] concentrates on elucidating the semantic analysis of those terms for which the logic has the resources: since Montague's intensional logic IL includes a time parameter, in depth analysis of temporal markers (tense, aspect, time adverbials) becomes possible. But as long as the logic lacks analogous resources for space, kinship terms, sensory inputs, or obligations, this approach has no traction, and heaping all these issues on top of what was already a computationally intractable logic calculus has not proven fruitful.

First Order Logic (FOL) is a continental divide in this regard. From a mathematical perspective, FOL is a small system, considering that the language of set theory requires only one binary relation, $\in$, and it is evident both from the Peano and the ZF axioms that you will need all well-formed formulas (or at least the fragment that has no atomic sentence lying in the scope of more than three quantifiers, see Tarski and Givant [41]) to do arithmetic. Therefore, those who believe that mathematics is but a small, clean, well-organized segment of natural language will search for the appropriate semantics somewhere upwards of FOL – this is the Montague Grammar (MG) tradition, where higher order intensional logic is viewed as essential. There is already significant work in trying to restrict the power of the Turing-complete higher order intensional apparatus to FOL (Blackburn and Bos [5]) and here we take this further, moving to formalisms that fall at the low end of the complexity scale, well below FOL. At that point, much of what mathematical logic offers is not applicable, and methods of algebra have more traction, as will be discussed in Section 2 in more detail. It is widely accepted that "people who put knowledge into computers need mathematical logic, including quantifiers, as much as engineers need calculus" (McCarthy [32]) but we claim that these tools are neither available in natural language (as noted repeatedly by the inventors of modern mathematical logic from Frege and Russell to Tarski) nor are they required for the analysis of natural language text – in the MG-style analysis it is the needs of the computer programmer that are being catered to at the expense of modeling the actual cognitive capabilities of the native speaker. This is not to say that such needs, especially for the engineer building knowledge-based systems, are not real, but our thesis is that the formalism appropriate for natural language semantics is too weak to supply this, being capable of natively supporting only a far weaker form of analogical reasoning discussed in Section 4.

In this paper we offer a formal theory of lexical definitions. A *word* that is to be defined will be given in italics; its definition will use for the most part unary atoms, given in `typewriter font` and to a lesser extent binary atoms, given is SMALL CAPS; its phonological representation (which we will also call its printname) will be marked by underscoring. Aside from the fancy typography, this is very much in keeping with linguistic tradition where a *sign* is conceived of as an ordered pair of `meaning` and form. (The typographical distinctions will pay off only in making the formal parts easier to parse visually – in running text, we will also use italics for emphasis and for the introduction of technical terms.) While we will have little to say about pronunciation, paradigmatic forms, style, or etimology here, the fact that these are important to the practice of lexicography is always kept in mind, and we will make an effort to indicate, however programmatically, how these are to be subsumed under the overall theory presented here.

Given the widely accepted role of the lexicon in grammatical theory as the storage place of last resort, containing all that is idiosyncratic, arbitrary, and language-particular, the question must be asked: why should anyone want to dive in this trashcan? First, we need to see clearly that the lexicon is not trash, but rather it is the essential fuel of all communicative effort. As anyone trying to communicate in a language they mastered only at a tourist level will know, lack of crisp grammar is rarely a huge barrier to understanding. If you can produce the words, native speakers will generally be forgiving if the conjugation is shaky or the proper auxiliary is missing. But if you don't have the words for beef stew or watch repairman, knowing that the analytic present perfect combines stage-level and individual-level predication and thus gives rise to an inchoative meaning will get you nowhere.

A more rigorous estimate of the information content of sentences confirms our everyday experience. The word entropy of natural language is about 12-16 bits/word (see Kornai [26]:7.1 for how this depends on the language in question). The number of binary parse trees over n nodes is $C_n \sim 4^n/\sqrt{\pi}n^{1.5}$ or less than 2 bits per word. Aronoff[4] describes in some detail how the Masoretes used only 2 bits (four levels of symbols) to provide a binary parse tree for nearly every Biblical verse – what we learned of coding since would now enable us to create an equally sparse system that is sufficiently detailed to cover *every* possible branching structure with slightly *less than* two bits on the average. Definitions of logical structure other than by parse tree are possible, but they do not alter the picture significantly: logical structure accounts for no more than 12-16% of the information conveyed by a sentence, a number that actually goes down with increased sentence length.

Another equally important reason why we need to develop a formal theory of word meaning is that without such a theory it is impossible to treat logical arguments like *God cannot create a mountain without creating a valley* which are based on the meaning of the predicates rather than on the meaning of the logical connectives. Why is this argument correct, even if we assume an omnipotent God? Because *mountain* means something like `land higher than surrounding` `land` so for there to be a mountain there needs to be a lower reference land, if there was no such reference ‘valley’ the purported mountain wouldn’t actually be a mountain. For St. Thomas Aquinas the argument serves to demonstrate that even God is bound by the laws of logic, and for us it serves as a reminder that the entire Western philosophical tradition from Aristotle to the Schoolmen considered word meaning an essential part of logic. We should add here that the same is true of the Eastern tradition, starting with Confucius’ theory of *cheng ming* (rectification of names) – for example, one who rules by force, rather than by the decree of heaven, is a *tyrant*, not a *king* (see Graham [16]:29). Modern mathematical logic, starting with De Morgan, could succeed in identifying a formal framework that can serve as a foundation of mathematics without taking the meaning of the basic elements into account because mathematical content differs from natural language content precisely in being lodged in the axioms entirely. However, for machine understanding of natural language text, lacking a proper theory of the meaning of words is far more of a bottleneck than the lack of compositional semantics, as McCarthy [31], and the closely related work on naive physics (Hayes [18]) already made clear.

What does a theory of the lexicon have to provide? First, adequate support for the traditional lexicographic tasks such as distinguishing word senses, deciding whether two words/senses are synonymous or perhaps antonymous, whether one expression can be said to be a paraphrase of another, etc. Second, it needs to connect to a theory of the meaning of larger (non-lexicalized) constructions including, but not necessarily limited to, sentential syntax and semantics. Third, it should provide a means of linking up meanings across languages, serving as a translation pivot. Fourth, it should be coupled to some theory of inference that enables, at the very least, common sense reasoning about objects, people, and natural phenomena. Finally, the theory should offer learning algorithms whereby the representation of meanings can be acquired by the language learner.

In this paper we disown the problem of *learning*, how an English-speaking child associates *water* with the sensory input (see Keller [21]), as it belongs more in cognitive science and experimental psychology than in mathematical linguistics, and the problem of *pattern recognition:* when is a person *fat*? It is possible to define this as the outcome of some physical measurements such as the Body Mass Index, but we will argue at some length that this is quite misguided. This is not to say that there is no learning problem or pattern recognition problem, but before we can get to these we first need a theory of what to learn and recognize.

This is not the place to survey the history of lexical semantics, and we confine ourselves to numerical estimates of coverage on the core vocabulary. The large body of analytic work on function words such as connectives, modals, temporals, numerals, and quantifiers covers less than 5% of core vocabulary, where 90% are content words. Erring on the side of optimism and assuming that categories of space, case in particular, can be treated similarly, would bring this number up to 6%, but not further, since the remaining large classes of function words, in particular gender and class markers, are clearly non-logical. Another large body of research approaches natural kinds by means of species and genera. But in spite of its venerable roots, starting with Aristotle's work on *eidopoios diaphora*, and its current popularity, including WordNet, EuroWordNet, and AsiaWordNet on the one hand and Semantic Web description logic (OWL) on the other, this method covers less than 10% of core vocabulary. This is still a big step forward in that it is imposing a formal theory on some content words, by means of a technique, default inheritance along IS_A links, that is missing from standard logic, including the high-powered modal intensional logics commonly used in sentential semantics. Perhaps surprisingly, the modern work on verb classification including Gruber [17], Dowty [9], Levin [29], FrameNet (Fillmore [12]), and VerbNet (Kipper et al [24]) has far broader scope, covering about 25% of core vocabulary.

Taking all these together, and assuming rather generously that all formal problems concerning these systems have been resolved, this is considerably less than half of the core vocabulary, and when it comes to the operations on these elements, all the classical and modern work on the semantics associated with morphological operations (Pāṇini, Jakobson, Kiparsky) covers numerically no more than 5-10% of the core operations. That the pickings of the formal theory are rather slim is especially clear if we compare its coverage to that of the less formally stated, but often strikingly insightful work in linguistic semantics, in particular to the work of Wierzbicka, Lakoff, Fauconnier, Langacker, Talmy, Jackendoff, and others often broadly grouped together as 'cognitively inspired'. We believe that part of the reason why the formal theory has so little traction is that it aims too high, largely in response to the well-articulated needs of AI and KR.

2 The Basic Elements

In creating a formal model of the lexicon the key difficulty is the circularity of traditional dictionary definitions – the first English dictionary, Cawdrey [7] already defines *heathen* as `gentile` and *gentile* as `heathen`. The problem has already been noted by Leibniz (quoted in Wierzbicka [45]):

> Suppose I make you a gift of a large sum of money saying you can collect it from Titius; Titius sends you to Caius; and Caius, to Maevius; if you continue to be sent like this from one person to another you will never receive anything.

One way out of this problem is to come up with a small list of primitives, and define everything else in terms of these. There are many efforts in this direction (the early history of the subject is discussed in depth in Eco [10]) but the modern efforts begin with Ogden's [35] Basic English. The KR tradition begins with the list of primitives introduced by Schank [40], and a more linguistically inspired list is developed by Wierzbicka and the NSM school. But it is not at all clear how Schank or Wierzbicka would set about defining new words based on their lists (the reader familiar with their systems should try to apply them to any term that is not on their lists such as *liver*). As a result, in cognitive science many have practically given up on meaning decomposition as hopeless. For example Mitchell et al. [33] distinguish words from one another by measuring correlation with their core words in the Google 5-gram data. Such correlations certainly do not constitute a semantic representation in the deductive sense we are interested in, but it requires no artful analysis, indeed, it requires no human labor at all, to come up with numerical values for any new word.

Here we sketch a more systematic approach that exploits preexisting lexicographic work, in particular dictionary definitions that are already restricted to a smaller wordlist such as the Longman Defining Vocabulary (LDV) or Ogden's Basic English (BE). These already have the proven capability to define all other words in the Longman Dictionary of Contemporary English (LDOCE) or the Simple English wikipedia at least for human readers, though not necessarily in sufficient detail and precision for reasoning by a machine. Any defining vocabulary **D** subdivides the problem of defining the meaning of (English) words in two. First, the definition of other vocabulary elements in terms of **D**, which is our focus of interest, and second, defining **D** itself, based perhaps on primary (sensory) data or perhaps on some deeper scientific understanding of the primitives. A complete solution to the dictionary definition problem must go beyond a mere listing **D** of the defining vocabulary elements: we need both a formal model of each element and a specification of lexical syntax, which regulates how elements of **D** combine with each other (and possibly with other, already defined, elements) in the definition of new words.

We emphasize that our goal is to provide an *algebra* of lexicography rather than a generative lexicon (Flickinger [15], Pustejovsky [36]) of the sort familiar from generative morphology. A purely generative approach would start from some primitives and some rules or constraints which, when applied recursively, provide an algorithm that enumerates the lexicon. The algebraic approach is more modest in that it largely leaves open the actual contents of the lexicon. Consider the semantics of noun-noun compounds. As Kiparsky [22] notes, *ropeladder* is 'ladder *made of* rope'; *manslaughter* is 'slaughter *undergone by* man'; and *testtube* is 'tube *used for* test', so the overall semantics can only specify that N_1N_2 is 'N_2 that is V-ed by N_1', i.e. the decomposition is subdirect (yields a superset of the target) rather than direct, as it would be in a fully compositional generative system.

Another difference between the generative and the algebraic approach is that only the former implies commitment to a specific set of primitives. To the extent that work on lexical semantics often gets bogged down in a quest for the ultimate primitives, this point is worth a small illustrative example. Consider the

Table 1. Multiplication in Z_3

	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

cyclic group Z_3 on three points given by the elements e, a, b and the preceding multiplication table.

The unit element e is unique (being the one and only y satisfying $yx = xy = x$ for all x) but not necessarily irreducible in that if a and b are given, both ab and ba could be used to define it. Furthermore, if a is given, there is no need for b in that aa already defines this element, so the group can be presented simply as $a, aa, aaa = e$ i.e. a is the 'generator' and $a^3 = e$ is the 'defining relation' (as these terms are used in group theory). Note, however, that the exact same group is equally well presented by using b as the generator and $b^3 = e$ as the defining relation – there is no unique/distinguished primitive as such. This non-uniqueness is worth keeping in mind when we discuss possible defining vocabularies.

In algebra, similar examples abound: for example in a linear space any basis is just as good as any other to define all vectors in the space. For a lexical example, consider the Hungarian verbal stem *toj* and the derived *tojó* 'hen', *tojás* 'egg', and *tojni* 'to lay egg'. It is evident that eggs are what hens lay, hens are what lay eggs, and laying of eggs is what hens do. In Hungarian, the interdependence of the definitions is made clear by the fact that all three forms are derived from the same stem by productive processes, *-ó* is a noun-forming deverbal suffix denoting the agent, *-ás* denotes the action or the result, and *-ni* is the infinitival suffix. But the same arbitrariness in the choice of primitives can be just as evident in less transparent examples, where the common stem is lacking: for example in English *hen* and *egg* it is quite unclear which one is logically prior. Consider *prison* 'place where inmates are kept by guards', *guard* 'person who keeps inmates in prison', and *inmate* 'person who is kept in prison by guards'. One could easily imagine a language where prison guards are called *keepers,* inmates *keepees*, and the prison itself a *keep*. The mere fact that in English the semantic relationship is not signaled by the morphology does not mean that it's not there – to the contrary, we consider it an accident of history, beyond the reach of explanatory theory, that the current nominal sense of *keep,* 'fortress' is `fortified place to keep the enemy out` rather than `to keep prisoners in`.

What is, then, a reasonable defining vocabulary **D**? We propose to define one from the outside in, by analyzing the LDV or BE rather than building from the inside out from the putative core lists of Schank or Wierzbicka. This method guarantees that at any given point of reducing **D** to some smaller **D'** we remain capable of defining all other words, not just those listed in LDOCE (some 90k items) or the Simple English wikipedia (over 30k entries) but also those that are definable in terms of these larger lists (really, the entire unabridged vocabulary of English). In the computational work that fuels the theoretical analysis presented here we begin with our own version of the LDV, called 4lang, which includes Latin, Hungarian, and Polish translations in the intended senses, both because we do not wish to lose sight of the longer term goal of translation and as a clear means of disambiguation for concepts whose common semantic root, if there ever was one, is no longer transparent, e.g. *interest* 'usura' v. *interest* 'studium'.

Clearly, a similarly disambiguated version of the BE vocabulary, or any other reasonable starting point could just as well be used.

We perform the analysis of the starting **D** in several chunks, many corresponding to what old-fashioned lexicographers would call a *semantic field* (Trier [42]), conceptually related terms that are likely candidates to be defined in terms of one another such as color terms, legal terms, and so on. We will not attempt to define the notion of semantic fields in a rigorous fashion, but use an operational definition based on Roget's Thesaurus. For example, for *color* terms we take about 30 stanzas from Roget 420 Light to Roget 449 Disappearance, (numbering follows the 1911 edition of Roget's as this is available as a Project Gutenberg etext #10681) and for *religious* terms we take 25 stanzas Roget 976 Deity to Roget 1000 Temple. Since the chunking is purely pragmatic, we need not worry about the issues that plague semantic fields: for our purposes it matters but little where the limits of each field are, whether the resulting collections of words and concepts are properly named, or whether some kind of hierarchy can or should be imposed on them – all that matters is that each form a reasonable unit of workable size, perhaps a few dozen to a few hundred stanzas. We will mostly use the Religion field to illustrate our approach, not because we see it as somehow privileged but rather because it serves as a strong reminder of the inadequacy of the physicalist approach. In discussing color, we may be tempted to dispense with a defining vocabulary **D** in favor of a more scientifically defined core vocabulary, but in general such core expressions, if truly restricted to measurable qualia, have very limited traction over much of human social activity.

The main fields defined through Roget are *size* R031 – R040a and R192 – R223; *econ* R775 – R819; *emotion/attitude* R820 – R936 except 845-852 and 922-927; *esthetics* R845 – R852; *law/morals* R937 – R975 plus R922 – 927. In this process, about a quarter of the LDV remains unaffiliated. For Religion we obtain the list *anoint, believe, bless, buddhism, buddhist, call, ceremony, charm, christian, christianity, christmas, church, clerk, collect, consecrated, cross, cure, devil, dip, doubt, duty, elder, elect, entrance, fairy, faith, faithful, familiar, fast, father, feast, fold, form, glory, god, goddess, grace, heaven, hinduism, holy, host, humble, jew, kneel, lay, lord, magic, magician, mass, minister, mosque, move, office, people, praise, pray, prayer, preserve, priest, pure, religion, religious, reverence, revile, rod, save, see, service, shade, shadow, solemn, sound, spell, spirit, sprinkle, temple, translate, unity, word, worship.* (Entries are lowercased for ease of automated stemming etc.)

Two problems are evident from such a list. First, there are several words that do not fully belong in the semantic field, in that the sense presented in Roget's is different from the sense in the LDV: for example *port* is not a color term and *father* is not a religious term in the primary sense used in the LDV. Such words are manually removed, since defining the religious sense of *father* or the color sense of *port* would in no way advance the cause of reducing the size of **D**. Programmatic removal is not feasible at this stage: to see what the senses are, and thus to see that the core sense is not the one used in the field, would require a working theory of lexical semantics of the sort we are developing here. Once such a theory is at hand, we may use it to verify the manual work performed early on, but this is only a form of error checking, rather than learning something new about the domain. Needless to say, *father* still needs to be defined or declared a primitive, but the place to do this is among kinship terms not religious terms.

If a word is kept, this does not mean that it is unavailable outside the semantic field, clearly *Bob worships the ground Alice walks on* does not mean anything religious. However, for words inside the field such as *worship* even usage external to the field relies on the field-internal metaphor, so the core/defining sense of the word is the one inside. Conversely, if usage does not require the field-internal metaphor, the word/sense need not be treated as part of the size reduction effort: for example, *This book fathered a new genre* does not mean (or imply) that the object will treat the subject with reverence, so *father* can be left out of the *religion* field. Ideally, with a full sense-tagged corpus one could see ways of making such decisions in an automated fashion, but in reality creating the corpus would require far more manual work than making the decisions manually.

Since the issue of different word senses will come up many times, some methodological remarks are in order. Kirsner [25] distinguishes two polarly opposed approaches. The *polysemic* approach aimed at maximally distinguishing as many senses as they appear distinct, e.g. *bachelor*$_1$ 'unmarried adult man', *bachelor*$_2$ 'fur seal without a mate', *bachelor*$_3$ 'knight serving under the banner of another knight', and *bachelor*$_4$ 'holder of a BA degree'. The *monosemic* approach (also called *Saussurean* and *Columbia School* approach by Kirsner, who calls the polysemic approach *cognitive*) searches for a single, general, abstract meaning, and would subsume at least the first three senses above in a single definition, 'unfulfilled in typical male role'. This is not the place to fully compare and contrast the two approaches (Kirsner's work offers an excellent starting point), but we note here a significant advantage of the monosemic approach, namely that it makes interesting predictions about novel usage, while the predictions of the polysemic approach border on the trivial. To stay with the example, it is possible to envision novel usage of *bachelor* to denote a contestant in a game who wins by default (because no opponent could be found in the same weight class or the opponent was a no-show). The polysemic theory would predict that not just seals but maybe also penguins without a mate may be termed *bachelor* - true but not very revealing.

The choice between monosemic and polysemic analysis need not be made on a priori grounds: even the strictest adherent of the polysemic approach would grant that *bachelor's degree* refers, at least historically, to the same kind of apprenticeship as *bachelor knight*. Conversely, even the strictest adherent of the monosemic approach must admit that the relationship between 'obtaining a BA degree' and 'being unfulfilled in a male role' is no longer apparent to contemporary language learners. That said, we still give methodological priority to the monosemic approach because of the original Saussurean motivation: if a single form is used, the burden of proof is on those who wish to posit separate meanings (see Ruhl [39]). An important consequence of this methodological stance is that we will rarely speak of *metaphorical* usage, assuming instead that the core meaning already extends to such cases.

A second problem, which has notable impact on the structure of the list, is the treatment of natural kinds. By natural kinds here we mean not just biologically defined kinds as *ox* or *yak*, but also culturally defined artifact types like *tuxedo* or *microscope* – as a matter of fact the cultural definition has priority over the scientific definition when the two are in conflict. The biggest reason for the inclusion of natural kinds in the LDV is not conceptual structure but rather the eurocentric viewpoint of LDOCE: for the English speaker it is reasonable to define the yak as ox-like, but for a Tibetan defining the ox as yak-like would make more sense. There is nothing wrong with being eurocentric in a dictionary of an Indoeuropean language, but for our purposes neither of these terms can be truly treated as primitive.

So far we discussed the *lexicon*, the repository of linguistic knowledge about words. Here we must say a few words about the *encyclopedia*, the repository of world knowledge. While our goal is to create a formal theory of lexical definitions, it must be acknowledged that such definitions can often elude the grasp of the linguist and slide into a description of world knowledge of various sorts. Lexicographic practice acknowledges this fact by providing, somewhat begrudgingly, little pictures of flora, fauna, or plumbers' tools. A well-known method of avoiding the shame of publishing a picture of the yak is to make reference to `Bos grunniens` and thereby point the dictionary user explicitly to some encyclopedia where better information can be found. We will collect such pointers in a set **E**, and use curly braces to set them typographically apart from references to lexical content.

When we say that *light* is defined as {`flux of photons in the visible band`}, what this really means is that `light` must be treated as a primitive. There is a physical theory of light which involves photons, a biophysical theory of visual perception that involves sensitivity of the retina to photons of specific wavelengths, but we are not interested in these theories, we are just offering a pointer to the person who is. From the linguistic standpoint *light* is a primitive, irreducible concept, one that people have used for millennia before the physical theory of electromagnetic radiation, or even the very notion of photons, was available. Ultimately any system of definitions must be rooted in primitives, and we believe the notion `light` is a good candidate for such a primitive. From the standpoint of lexicography only two things need to be said: first, whether we intend to take the nominal or the verbal meaning as our primitive, and second, whether we believe that the primitive notion `light` is shared across the oppositions with dark and with heavy or whether we have two different senses of *light*. In this particular case, we choose the second solution, treating the polysemy as an accident of English rather than a sign of deep semantic relationship, but the issue must be confronted every time we designate an element as primitive. The issue of how to assign grammatical category (also called part of speech or POS) to the primitives will be discussed in Section 3, but we note here in advance that we keep the semantic part of the representation constant across verbs, their substantive forms, and their cognate objects.

The same point needs to be made in regards to ontological primitives like *time.* While it is true that the time used in the naive physics model is discrete and asynchronous, this is not intended as some hypothesis concerning the ultimate truth about physical time, which appears continuous (except possibly at a Planck scale) and appears distinct from space and matter (but is strongly intertwined with these). We take the appropriate method for deciding such matters to be physical experimentation and theory-making, and we certainly do not propose to find out the truth of the matter by reverse-engineering the lexica of natural languages. Since the model is not intended as a technical tool for the analysis of synchrony or continuous time, we do not wish to burden it with the kind of mechanisms, such as Petri nets or real numbers, that one would need to analyze such matters. Encyclopedic knowledge of `time` may of course include reference to the real numbers or other notions of continuous time, but our focus is not with a deep understanding of time as with tense marking in natural language, and it is the grammatical model, not the ontology, that carries the burden of recapitulating this. For the sake of concreteness we will assume a Reichenbachian view, distinguishing four different notions of time: (i) *speech time,* when the utterance is spoken, (ii) *perspective time,* the vantage point of temporal deixis, (iii) *reference time,* the time that adverbs refer to, and (iv) *event time,* the time the named event unfolds. Typically, these are intervals, possibly open-ended, more rarely points (degenerate intervals) and the hope is that we can eventually express the temporal semantics of natural language in terms of interval relations such as 'event time precedes reference time' (see Allen [1], [2], Kiparsky [23]). The formal apparatus required for this is considerably weaker than that of FOL.

One important use of external pointers worth separate mention is for proper names. By *sun* we mean primarily the star nearest to us. The common noun usage is secondary, as is clear from the historical fact that people before Giordano Bruno didn't even know that the small points of light visible on the night sky were also suns. That we have a theory of the Sun as {`the nearest star`} where `the, near, -est,` and `star` are all members of the LDV is irrelevant from a lexicographic standpoint – what really matters is that there is a particular object, ultimately identified by deixis, that is a natural kind on its own right. The same goes for natural kinds such as *oxygen* or *bacteria* that may not even have a naive lexical theory (it is fair to say that all our knowledge about these belongs in chemistry and the life sciences) and about cultural kinds such as *tennis, television, british,* or *october.* In 3.3 we return to the issue of how to formalize those cases when purely lexical knowledge is associated with natural kinds, e.g. that tennis is a game played with a ball and rackets, that November follows October, or that bacteria are small living things that can cause disease, but we wish to emphasize at the outset that there is much in the encyclopedia that our formalism is not intended to cover, e.g. that the standard atomic weight of oxygen is 15.9994(3). Lest the reader feel that any reference to some external encyclopedia is tantamount to shirking of lexicographic duty it is worth keeping in mind that natural and cultural kinds amount to less than 6% of the LDV.

Returning to the field of religion, when we define *Islam* as `religion centered on the teachings of {Mohamed}`, the curly braces acknowledge the fact Mohamed (and similarly Buddha, Moses, or Jesus Christ) will be indispensable in any effort aimed at defining Islam (Buddhism, Judaism, or Christianity). The same is true for Hinduism, which we may define as being centered on `revealed teachings ({śruti})`, but of course to obtain Hinduism as the definiendum the definiens must make it clear that it is not any old set of revealed teachings that are central to it but rather the Vedas and the Upanishads. One way or another, when we wish to define such concepts as specific religions, some reference to specific people and texts designated by proper names is unavoidable. Remarkably, once the names of major religious figures and the titles of sacred texts are treated as pointers to the encyclopedia, there remains nothing in the whole semantic field that is not definable in terms of non-religious primitives. In particular, *god* can be defined as `being, supreme` where `supreme` is simply about occupying the highest position in a hierarchy (being a `being` has various implications, see Section 3.1, but none of these are particularly religious). The same does not hold for the semantic field of color, where we find irreducible entries such as `light`.

Needless to say, our interest is not with exegesis (no doubt theologians could easily find fault with the particular definitions of god and the major religions offered here) but with the more mundane aspects of lexicography. Once we have *buddhism, christianity, hinduism, islam,* and *judaism* defined, *buddhist, christian, hindu, muslim,* and *jew* fall out as `adherent of buddhism, ..., judaism` for the noun denoting a person, and similarly for the adjectives *buddhist, christian, hindu, islamic, jewish* which get defined as `of or about buddhism,..., judaism`. We are less concerned with the theological correctness of our definitions than with the proper choice of the base element: should we take the *-ism* as basic and the *-ist* as derived, should we proceed the other way round, or should we, perhaps, derive both (or, if the adjectival form is also admitted, all three) from a common root? Our general rule is to try to derive the morphologically complex from the morphologically simplex, but exceptions must be made e.g. when we treat *jew* as derived (as if the word was **judaist*). These are well handled by some principle of blocking (Aronoff [3]), which makes the non-derived jew act as the printname for **judaist.*

Another, seemingly mundane, but in fact rather thorny issue is the treatment of bound morphemes. The LDV includes, with good reason, some forty suffixes *-able, -al, -an, -ance, -ar, -ate, -ation, -dom, -en, -ence, -er, -ess, -est, -ful, -hood, -ible, -ic, -ical, -ing, -ion, -ish, -ist, -ity, -ive, -ization, -ize, -less, -like, -ly, -ment, -ness, -or, -ous, -ry, -ship, -th, -ure, -ward, -wards, -work, -y* and a dozen prefixes *counter-, dis-, en-, fore-, im-, in-, ir-, mid-, mis-, non-, re-, un-, vice-, well-*. This affords great reduction in the size of **D**, in that a stem such as *avoid* now can appear in the definiens in many convenient forms such as `avoidable, avoidance, avoiding` as the syntax of the definition dictates. Including affixes is also the right decision from a cross-linguistic perspective, as it is evident that notions that are expressed by free morphemes in one language, such as possession

(English *my, your, ...*), are expressed in many other languages by affixation. But polysemy can be present in affixes as well: for example, English and Latin have four affixes *-an/anus, -ic/ius, -ical/icus,* and *-ly/tus* where Hungarian and Polish have only one *-i/anin* and we have to make sure that no ambiguity is created in the definitions by the use of polysemous affixes. Altogether, affixes and affix-like function words make up about 8-9% of the LDV, and the challenge they pose to the theory developed here is far more significant than that posed by natural kinds in that their proper analysis involves very little, if any, reference to encyclopedic knowledge.

Finally, there is the issue of the economy afforded by primitive conceptual elements that have no clear exponent in the LDV. For example, we may decide that we feel *sorrow* when something bad happens to us, *gloating* when it happens to others, *happiness* when something good happens to us, and *resentment* when it happens to others. (The example is from Hobbs [19], and there is no claim here or in the original that these are the best or most adequate emotional responses. Even if we agree that they are not, this does not affect the following point, which is about the economy of the system rather than about morally correct behavior.) Given that `good, bad`, and `happen` are primitives we will need in many corners of the system, we may wish to rely on some sociological notion of `in-group` and `out-group` rather than on the pronouns `us` and `them` in formalizing the above definitions. This has the clear advantage of remaining applicable independent of the choice of in-group (be it family, tribe, nation, colleagues, etc) and of indexical perspective (be it ours or theirs). Considerations of economy dictate that we use abstract elements as long as we can reduce the defining vocabulary **D** by more than one item: whether we prefer to use `in-group, out-group` or `us, them` as primitives is more a matter of taste than a substantive issue. If two solutions **D** and **D'** have the same size, we have no substantive reason to prefer one to the other. That said, for expository convenience we will still prefer non-technical to technical and Anglo-Saxon to latinate vocabulary in our choice of primitives.

To summarize what we have so far, for the sake of concreteness we identified a somewhat reduced version of the LDV, less than 2,000 items, including some bound morphemes and natural kinds, as our defining vocabulary **D**, but we make no claim that this is in any way superior to some other base list **D'** as long as **D'** is not bigger than **D**.

3 The Formal Model

The key issue is not so much the membership of **D** as the mechanism that regulates how its elements are put together. Here we depart from the practice of the LDOCE, which uses natural language paraphrases, in favor of a fully formal theory. In 3.1 we introduce the elements of this theory which we will call *lexemes.* In 3.2 we turn to the issue of how these elements are combined with one another. The semantics of the representations is discussed in 3.3. The formalism is introduced gradually, establishing the intuitive meaning of the various components before the fully formal definitions are given.

3.1 Lexemes

We will call the basic building blocks of our system *lexemes* because they offer a formal reconstruction of the informal notion of lexicographic lexemes. Lexemes are well modularized knowledge containers, ideally suited for describing our knowledge of words (as opposed to our encyclopedic knowledge of the world, which involves a great deal of non-linguistic knowledge such as motor skills or perceptual inputs for which we lack words entirely). Lexemes come in two main varieties, *unary* lexemes which correspond to most nouns, adjectives, verbs, and content words in general (including most transitive and higher arity verbs as well) will be written in `typewriter font`, and *binary* lexemes, corresponding to adpositions, case markers, and other linkers, will be written SMALL CAPS. Ignoring the printnames, the *base* of unary lexemes consists of an unordered (conjunctive) list of properties, e.g. the *dog* is `four-legged, animal, hairy, barks, bites, faithful, inferior`; the *fox* is `four-legged, animal, hairy, red, clever`.

Binary lexemes are to be found only among the function words: for example AT(x,y) 'x is at location y', HAS(x,y) 'x possesses y', CAUSE(x,y) etc. In what follows these will be written infix, which lets us do away with variables entirely. (Thus the notation already assumes that there are no true ditransitives, a position justified in more detail in Kornai [27].) Binary lexemes have two defining lists of properties, one list pertaining to their first (superordinate) argument and another to their second (subordinate) argument – these two are called the *base* of the lexeme. We illustrate this on the predicate HAS, which could be the model for verbs such as *owns, has, possesses, rules,* etc. The differences between John HAS Rover and Rover HAS John are best seen in the implications (defaults) associated with the superordinate (possessor) and subordinate (possessed) slots: the former is assumed to be independent of the latter, the latter is assumed to be dependent on the former, the former controls the latter (and not the other way around), the former can end the possession relationship unilaterally, the latter can not, etc. The list of definitional properties is thus partitioned in two: those that belong to the superordinate argument are collected in the *head* partition, those belonging to the subordinate argument are listed on the *dependent* partition.

The lexical entries in question may also include pointers to sensory data, biological, visual, or other extralinguistic knowledge about dogs and foxes. We assume some set **E** of external pointers (which may even be two-way in the sense that external sensory data may trigger access to lexical content) to handle these, but here **E** will not be used for any purpose other than delineating linguistic from non-linguistic concerns. How about the defining elements that we collected in **D**? These are no different, their definitions can refer to other lexemes that correspond to their essential properties. So definitions can invoke other definitions, but the circularity causes no foundational problems, as argued above.

Following Quillian [37], semantic networks are generally defined in terms of some distinguished links: IS_A to encode facts such as dogs are animals, and ATTR to encode facts such that they are hairy. Here neither the genus nor the attribution relation is encoded explicitly. Rather, everything that appears on the distinguished (head) partition is attributed (or predicated) directly, and IS_A is defined simply by containment of the essential properties. Elementary pieces of link-tracing logic, such as A IS_A B ∧ B IS_A C ⇒ A IS_A C or A IS_A B ∧ B HAS C ⇒ A HAS C follow without any stipulation if we adopt this definition, but the system becomes more redundant: instead of listing only essential properties of dogs we need to list all the essential properties of the supercategories such as animals as well. Altogether, the use of IS_A links leads to better modularized knowledge bases, and for this reason we retain them as a presentation device, but without any special status: for us *dog* IS_A *animal* is just as valid as *dog* IS_A *hairy* and *dog* IS_A *barks*. From the KR perspective the main point here is that there is no mixing of strict and default inheritance, in fact there no strict portion of the system (except possibly in the encyclopedic part which need not concern us here).

If we know that animals are alive then we know that donkeys are alive. If we know that being alive implies life functions such as growth, metabolism, and replication this implication will again be inherited by animals and thus by mules as well. The encyclopedic knowledge that mules don't replicate has to be learned separately. Once acquired, this knowledge will override the default inheritance, but we are equally interested in the *naive* world-view where such knowledge has not yet been acquired. Only the naive lexical knowledge will be encoded by primitives directly: everything else must be given indirectly, by means of a pointer or set of pointers to encyclopedic knowledge. The most essential information that the lexicon has about *tennis* is that it is a *game*, all the world knowledge that we have about it, the court, the racket, the ball, the pert little skirts, and so forth, are stored in a non-lexical knowledge base. This is also clear from the evidence from word-formation: clearly *table tennis* is a kind of *tennis*, yet it requires no court, has a different racket, ball, and so forth. The clear distinction between essential (lexical) and accidental (encyclopedic) knowledge has broad implications for the contemporary practice of Knowledge Representation, exemplified by systems like CyC (Lenat and Guha [28]) or Mindpixel in that the current homogeneous knowledge bases need to be refactored, splitting out a small, lexical base that is entirely independent of domain.

The syntax of well-formed lexemes can be summarized in a Context-Free Grammar (V, Σ, R, S) as follows. The nonterminals V are the start symbol S, the binary relation symbols B, and the unary relation symbols collected in U. Variables ranging over V will be taken from the end of the Latin alphabet, v, w, x, y, z. The terminals are the grouping brackets '[' and ']', the derivation history parentheses '(' and ')', and we introduce a special terminating operator ';' to form a terminal *v;* from any nonterminal v. The rule $S \rightarrow U|B|\lambda$ handles the decision to use unary or binary lexemes, or perhaps none at all. The operation of *attribution* is captured in the rule schema $w \rightarrow w;[S^*]$ which produces the list defining w. This requires the CFG to be *extended* in the usual sense that regular expressions are permitted on the right hand side, so the rule really means $w \rightarrow w;[]|w;[S]|w;[SS]|...$ Finally, the operation of *predication* is handled by $u \rightarrow u;(S)$ for unary, and $v \rightarrow Sv;S$ for binary nonterminals. All lexemes are built up recursively by these rules.

3.2 Combining the Lexemes

The first level of combining lexemes is morphological. At the very least, we need to account for productive derivational morphology, the prefixes and suffixes that are part of **D**, but in general we expect a theory that is just as capable of handling cases not easily exemplified in English such as binyanim. Compounding, to the extent predictable, also belongs here, and so does nominalization, especially as definitions make particularly heavy use of this process. The same is true for inflectional morphology, where the challenge is not so much English (though the core set *-s, 's, -ing, -ed* must be covered) as languages with more complex inflectional systems. Since certain categories (e.g. gender and class system) can be derivational in one language but inflectional in another, what we really require is *coverage of all productive morphology.* This is obviously a tall order, and within the confines of this paper all we can do is to discuss one example, deriving *insecure*, from *in-* and *secure*, as this will bring many of the characteristic features of the system in play.

Irrespective of whether *secure* is primitive (we assume it is not), we need some mechanism that takes the *in-* lexeme, the *secure* lexeme, and creates an *insecure* lexeme whose definition and printname are derived from those of the inputs. To forestall confusion we note here that not every morphologically complex word will be treated as derived. For example, it is clear, e.g. from the strong verb pattern, that *withstand* is morphologically complex, derived from *with* and *stand* (otherwise we would expect the past tense to be **withstanded* rather than *withstood*), yet we do not attempt to describe the operation that creates it. We are content with listing *withstand, understand,* and other complex forms in the lexicon, though not necessarily as part of **D**. Similarly, if we have a model capable of accounting for *insecure* in terms of more primitive elements, we are not required to overapply the technique to *inscrutable* or *ineffable* just because these words are also morphologically complex and could well be, historically, the residue of *in-* prefixation to stems no longer preserved in the language. Our goal is to define meanings, and the structural decomposition of every lexeme to irreducible units is pursued only to the extent it advances this goal.

Returning to *insecure,* the following facts should be noted. First, that the operation resides entirely in *in-* because *secure* is a free form. Second, that a great deal of the analysis is best formulated with reference to lexical categories (parts of speech): for example, *in-* clearly selects for an adjectival base and yields an adjectival output (the category of *in-* is A/A), because those forms such as *income* or *indeed* that are formed from a verbal or nominal base lack the negative meaning of *in-* that we are concerned with (and are clearly related to the preposition *in* rather than the prefix *in/im* that is our target here). Third, that the meaning of the operation is exhaustively characterized by the negation: forms like *infirm* where the base *firm* no longer carries the requisite meaning still carry a clear negative connotation (in this case, 'lacking in health' rather than 'lacking in firmness'). In fact, whatever meaning representation we assign to the lexically listed element *insecure* must also be available for the non-lexical (syntactically derived) *not secure.*

In much of model-theoretic semantics (the major exception is the work of Turner [43], [44]) preserving the semantic unity of stems like *secure* which can be a verb or an adjective, or stems like *divorce* which can be both nouns and verbs, with no perceptible meaning difference between the two, is extremely hard because of the differences in signature. Here it is clear that the verb is derived from the adjective: clearly, the verb *to secure x* means 'make x (be) secure', so when we say that *in-* selects for an adjectival base, this just means the part of the POS structure of *secure* that permits verbal combinatorics is filtered out by application of the prefix. The adjective *secure* means 'able to withstand attack'. Prefixation of *in-* is simply the addition of the primitive `neg` to the semantic representation and concatenation plus assimilation in the first, cf. *in+secure* and *im+precise*. (We note here, without going into details, that the phonological changes triggered by the concatenation are also entirely amenable to treatment in finite state terms.)

As far as the invisible deadjectival verb-forming affix (paraphrased as *make*) that we posited here to obtain the verbal form, this does two things: first, it brings a subject slot x, and second, it contributes a change of state predicate – before, there wasn't an object y, and now there is. The first effect, which requires making a distinction between an external (subject) and internal (direct object, indirect object, etc) arguments, follows a long tradition of syntactic analysis going back at least to Williams [46], and will just be assumed without argumentation here, but the latter is worth discussing in greater detail, as it involves a key operation among lexemes, *substitution*, to which we turn now.

Some form of recursive substitution of definitions in one another is necessary both for work aimed at reducing the size of the DV and for attempts to define non-**D** elements in terms of the primitives listed in **D**. When we add an element of negation (here given simply as `neg`, and a reasonable candidate for inclusion in **D**) to a definition such as 'able to withstand attack', how do we know that the result is 'not able to withstand attack' rather than 'able to not withstand attack' or even 'able to withstand not attack'? The question is particularly acute because the head just contains the defining properties as elements of a set, with no order imposed. (We note that this is a restriction that we could trivially give up in favor of ordered lists, but only at a great price: once ordered lists are admitted the system would become Turing-complete just as HPSG.) Another way of asking the same question is to ask how the system deals with iterated substitutions, for even if we assume that ABLE and `attack` are primitives (they are listed in the LDV), surely *withstand* is not, *x withstands y* means something like 'x does not change from y' or even 'x actively opposes y'. Given our preference for a monosemic analysis we take the second of these as our definition, but this makes the problem even more acute: how do we know that the negation does not attach to the *actively* portion of the definition? What is at stake here is the single most important property of definitions, that the definiens can be substituted for the definiendum in any context.

Since many processes, such as making a common noun definite, which are performed by syntactic means in English, will be performed by inflectional means in other languages such as Rumanian, *complete coverage of productive morphology in the world's languages* already implies coverage of a great deal of syntax in English. Ideally, we would wish to take this further, requiring coverage of syntax as a whole, but we could be satisfied with slightly less, covering the meaning of syntactic constructions only to the extent they appear in dictionary definitions. Remarkably, almost all problem cases in syntax are already evident in this restricted domain, especially as we need to make sure that constructions and idioms are also covered. There are forms of grammar which assume all syntax to be a combination of constructions (Fillmore and Kay [13]), and the need to cover the semantics of these is already clear from the lexical domain: for example, a *mule* is `animal, cross between horses and donkeys, stubborn, ...` Clearly, a notion such as 'cross between horses and donkeys' is not a reasonable candidate for a primitive, so we need a mechanism for feeding back the semantics of nonce constructions into the lexicon.

This leaves only the totally non-lexicalized, purely grammatical part of syntax out of scope, cases such as topicalization and other manipulation of given/new structure, as dictionary definitions tend to avoid communicative dynamics. But with this important caveat we can state the requirement that lexical semantics cover not just the lexical, but also the syntactic combination of morphemes, words, and larger units.

3.3 The Semantics of Lexemes

Now that we have seen the basic elements (lexemes) and the basic mode of combination (attribution, modeled as listing in the base of a lexeme), the question will no doubt be asked: how is this different from Markerese (Lewis [30])? The answer is that we will interpret our lexemes in model structures, and make the combination of lexemes correspond to operations on these structures, very much in the spirit of Montague [34]. Formally, we have a source algebra $\mathcal{A}$ that is freely generated from some set of primitives $\mathbf{D}$ by means of constructions listed in $\mathbf{C}$. An example of such a construction would be *x is to y as z is to w* which is used not just in arithmetic (proportions) but also in everyday analogy: *Paris is to London as France is to England*, but *in*-prefixation would also be a construction of its own. We will also have an algebra $\mathcal{M}$ of *machines*, which will serve as our model structures, and a mapping σ of semantic interpretation that will assign elements of $\mathcal{M}$ both to elements of $\mathbf{D}$ and to elements of $\mathcal{A}$ formed from these in a compositional manner. This can be restated even more compactly in terms of category theory: members of $\mathbf{D}$, plus all other elements of the lexicon, plus all expressions constructed from these, are the objects of some category L of linguistic expressions, whose arrows are given by the constructions and the definitional equations, members of $\mathcal{M}$, and the mappings between them, make up the category M, and semantic interpretation is simply a functor S from L to M.

The key observation, which bears repeating at this point, is that S *underdetermines* the semantics of lexicalized expressions: if noun-noun compounding (obviously a productive construction of English) has the semantics 'N_2 that is

V-ed by N_1' all the theory gives us is that *ropeladder* is a kind of ladder that has something to do with rope. What we obtain is `ladder, rope` rather than the desired `ladder, material, rope`. Regrettably, the theory can take us only so far – the rest has to be done by diving into the trashcan and cataloging historical accidents.

Lexemes will be mapped by S on finite state automata (FSA) that *act* on partitioned sets of elements of $\mathbf{D} \cup \underline{\mathbf{D}} \cup \mathbf{E}$ (the underlined forms are printnames). Each partition contains one or more elements of $\mathbf{D} \cup \mathbf{E}$ or the printname of the lexeme (which is, as a matter of fact, just another pointer, to phonetic/phonological knowledge, a domain that we happen to have a highly developed theory of). By *action* we mean a relational mapping, which can be one to many or many to one, not just permutation. These FSA, together with the mapping associating actions to elements of the alphabet, are *machines* in the standard algebraic sense (Eilenberg [11]), with one added twist: the underlying set, called the *base* of the machine, is *pointed* (one element of it is distinguished). The FSA is called the *control*, the distinguished point is called the *head* of the base.

Without control, a system composed of bases would be close to a semantic network, with activations flowing from nodes to nodes (Quillian [38]). Without a base, the control networks would just form one big FSA, a primitive kind of deduction system, so it is the combination of these two facets that give machines their added power and flexibility. Since the definitional burden is carried in the base, and the combinatorial burden in the control, the formal model has the resources to handle the occasional mismatch between syntactic type (part of speech) and semantic type (as defined by function-argument structure).

Let us now survey lexemes in order of increasing base complexity. If the base is empty, it has no relations, so the only FSA that can act on it is the null graph (no states and no transitions). This is called the NULL lexeme. If the set has one member, the only relations it can have is the identity 1 and the empty relation 0, which combine in the expected manner ($0{\cdot}0 = 0{\cdot}1 = 1{\cdot}0 = 0, 1{\cdot}1 = 1$). Note that the identity corresponds to the empty string usually denoted λ or ϵ. Since $1^n = 1$, the behavior of the machine can only take four forms, depending on whether it contains $0, 1$, both, or neither, the last case being indistinguishable from the NULL lexeme over any size base. If the behavior is given by the empty string alone, we will call the lexeme 1 with the usual abuse of notation, independent of the size of the base set. If the behavior is given by the empty relation alone, we will call the lexeme 0, again independent of the size of the base set. Slightly more complex is the lexeme that contains both 0 and 1, which is rightly thought of as the *union* of 0 and 1, giving us the first example of an operation on lexemes.

To fix the notation, in Table 2 we present the multiplication table of the semigroup R_2 that contains all relations over two elements (for ease of typesetting the rows and columns corresponding to 0 and 1 are omitted). The remaining elements are denoted $a, b, d, u, p, q, n, p', q', a', b', d', u', t$ – the prime is also used to denote an involution over the 16 elements which is *not* a semigroup

Table 2. Multiplication in R_2

	a	b	d	u	p	q	n	p’	q’	a’	b’	d’	u’	t
a	a	0	d	0	a	q	d	d	0	d	q	a	q	q
b	0	b	0	u	u	0	u	b	q’	q’	u	q’	b	q’
d	0	d	0	a	a	0	a	d	q	q	a	q	d	q
u	u	0	b	0	u	q’	b	b	0	b	q’	u	q’	q’
p	p	0	p’	0	p	t	p’	p’	0	p’	t	p	t	t
q	a	d	d	a	a	q	q	d	q	q	q	q	q	q
n	u	d	b	a	p	q’	1	p’	q	u’	d’	b’	a’	t
p’	0	p’	0	p	p	0	p	p’	t	t	p	t	p’	t
q’	u	b	b	u	u	q’	q’	b	q’	q’	q’	q’	q’	q’
a’	u	p’	b	p	p	q’	d’	p’	t	t	d’	t	a’	t
b’	p	d	p’	a	p	t	u’	p’	q	u’	t	b’	t	t
d’	p	b	p’	u	p	t	a’	p’	q’	a’	t	d’	t	t
u’	a	p’	d	p	p	q	b’	p’	t	t	b’	t	u’	t
t	p	p’	p’	p	p	t	t	p’	t	t	t	t	t	t

homomorphism (but does satisfy $x'' = x$). Under this mapping, $0' = t$ and $1' = n$, the rest follows from the naming conventions.

To specify an arbitrary lexeme over a two-element base we need to select an *alphabet* as a subset of these letters, an FSA that generates some language over (the semigroup closure of) this alphabet, and fix one of the two base elements as the head. (To bring this definition in harmony with the one provided by Eilenberg we would also need to specify input and output mappings α and ω but we omit this step here.) Because any string of alphabetic letters reduces to a single element according to the semigroup multiplication, the actual behavior of the FSA is given by selecting one of the 2^{16} subsets of the alphabet $[0, 1, a, \ldots, t]$, so over a two-element base there can be no more than 65,536, and in general over an n-element base no more than 2^{n^2} non-isomorphic lexemes, since over n elements there will be n^2 ordered pairs and thus 2^{n^2} relations. While in principle the number of nonisomorphic lexemes could grow faster than exponentially in n, in practice the base can be limited to three (one partition for the printname one for subject and one for object) so the largest lexeme we need to countenance will have its alphabet size limited to 512. This is still very large, but the upper bound is very crude in that not all conceivable relations over three elements will actually be used, there may be operators that affect subject and object properties at the same time but there aren’t any that directly mix grammatical and phonological properties.

Most nominals, adjectives, adadjectives, and verbs will only need one content partition. Relational primitives such as x AT y ‘x is at location y’; x HAS y ‘x is in possession of y’; x BEFORE y ‘x temporally precedes y’ will require two content partitions (plus a printname). As noted earlier, transitive and higher arity verbs will also generally require only *one* content partition: *eats*(x,y) may look superficially similar to *has*(x,y) but will receive a very different analysis. At this point, variables serve only as a convenient shorthand: as we shall see shortly, specifying the actual combinatorics of the elements does not require parentheses, variables, or an operation of variable binding. Formally we could use more complex lexemes for ditransitives like *give* or *show*, or verbs with even higher arity such as *rent,* but in practice we will treat these as combinations of primitives with smaller arity. e.g. *x gives y to z* as x CAUSE(z HAS y). (We will continue using both variables and natural language paraphrases as a convenient shorthand when this does not affect the argument we are making.)

Let us now turn to operations on lexemes. Given a set $\mathcal{L}$ of lexemes, each n-ary operation is a function from $\mathcal{L}^n$ to $\mathcal{L}$. As is usual, distinguished elements of $\mathcal{L}$ such as NULL, 0, and 1 are treated as nullary operations. The key unary operations we will consider are step, denoted ’; invstep, denoted ` ; and clean, denoted `-`. ’ is simply an elementary step of the FSA (performed on edges) which acts as a relation on the partition X. As a result of step R, the active state moves from x_0 to the image of x_0 under R. The inverse step does the opposite. The key binary operation is substitution, denoted by parens. The head of the dependent machine is built into the base of the head machine. For a simple illustration, recall the definition of *mule* as `animal, cross between horses and donkeys, stubborn,...` So far we said that one partition of the *mule* lexeme, the head, simply contains the conjunction (unordered list) of these and similar defining (essential) properties. Now assume, for the sake of the argument, that *animal* is not a primitive, but rather a similar conjunction `living, capable of locomotion,...` Substitution amounts to treating some part of the definiens as being a definiendum on its own right, and the substitution operation replaces the atomic `animal` on the list of essential properties defining *mule* by a conjunction `living, capable of locomotion,...` The internal bracketing is lost, what we have at the end of this step is simply a longer list `living, capable of locomotion, cross between horses and donkeys, stubborn,...`

By repeated substitution we may remove `living, stubborn`, etc. – the role of the primitives in **D** is to guarantee that this process will terminate. But note that the semantic value of the list is not changed if we leave the original `animal` in place: as long as animals are truly defined as living things capable of locomotion, we have set-theoretical identity between `animal, living, capable of locomotion` and `living, capable of locomotion` (cf. our second remark above). Adding or removing redundant combinations of properties makes no difference.

Let us now consider the next term, `cross between horses and donkeys`. By analyzing what `cross` means we can obtain statements father(donkey,mule) and mother(horse,mule). We will ignore all the encyclopedic details (such as the fact that if the donkey is female and the horse male the offspring is called a *hinny* not a mule) and concentrate on the syntax: how can we describe a statement such as

$$\forall x \text{ mule}(x)\ \exists y, z \text{ horse}(y) \& \text{ female}(y)\ \&\ \text{donkey}(z)\ \&\ \text{male}(z)\ \&\ \text{parent}(x, y)\ \&\ \text{parent}(x, z)$$

without recourse to variables? First, note that the Boolean connective & is entirely unnecessary, since everything is defined by a conjunction of properties – at best what is needed is to keep track of which parent has what gender, a matter that is generally handled by packing this information in a single lexical entry. Once we explain $\forall x$ mule(x) $\exists y$ horse(y) female(y) parent(x, y) the rest will be easy. Again, note that it makes no difference whether we consider a female horse or *mare* which is a parent or a horse which is a female parent or *mother*, these combinations will map out the exact same set. Whether primitives such as MOTHER, mare or being are available is a matter of how we design **D**.

Either way, further quantification will enter the picture as soon as we start to unravel *parent*, a notion defined (at least for this case) by 'gives genetic material to offspring' which in turn boils down to 'causes offspring to have genetic material'. Note that both the quantification and the identity of the genetic material are rather weak: we don't know whether the parent gives all its genetic material or just part of it, and we don't know whether the material is the same or just a copy. But for the actual definition none of these niceties matter: what matters is that mules have horse genes and donkey genes. As a matter of fact, this simple definition applies to hinnies as well, which is precisely the reason why people who lack significant encyclopedic knowledge about this matter don't keep the two apart, and even those who do will generally agree that a hinny is a kind of a mule, and not the other way around (just as bitches are a kind of a dog, i.e. the marked member of the opposition).

After all these substitution steps what remains on the list of essential mule properties includes complex properties such as HAS(horse genes) and capable of locomotion but no variable is required as long as we grant that in any definiens the superordinate (subject) slot of HAS is automatically filled by the definiendum. Readers familiar with the Accessibility Hierarchy of Keenan and Comrie [20] and subsequent work may jump to the conclusion that one way or another the entire hierarchy (handled in HPSG and related theories by an ordered list) will be necessary, but we attempt to keep the mechanism under much tighter control. In particular, we assume no ternary relations whatsoever, so there are no such things as indirect objects, let alone obliques, in definitions. To get further with capable of locomotion we need to provide at least a rudimentary theory of being capable of doing something, but here we feel justified in assuming that CAN, CHANGE, and PLACE are primitives, so that CAN(CHANGE(PLACE)) is good enough. Notice that what would have been the subject variables, who has the capability, who performs the change, and who has the place, are all implicitly bound to the same superordinate entity, the mule.

To make further progress on horse genes we also need a theory of compound nouns: what are horse genes if not genes characteristic of horses, and if they are indeed characteristic of horses how come that mules also have them, and in an essential fashion to boot? The key to understanding *horse gene* and similar compounds such as *gold bar* is that we need to supply a predicate that binds the two terms together, what classical grammar calls 'the genitive of material' that we will write as MADE_OF. A full analysis of this notion is beyond the limits of this paper, but we note that the central idea of MADE_OF is production, generation: the bar is produced from/of/by gold, and the genes in question are produced from/of/by horses. This turns the Kripkean idea of defining biological kinds by their genetic material on its head: what we assume is that horse genes are genes defined by their essential horse-ness rather than horses are animals defined by carrying the essence of horse-ness in their genes. (Mules are atypical in this respect, in that their essence can't be fully captured without reference to their mixed parentage.)

4 Conclusions

In the Introduction we listed some desiderata for a theory of the lexicon. First, adequate support for the traditional lexicographic tasks such as distinguishing word senses, deciding whether two words/senses are synonymous or perhaps antonymous, whether one expression can be said to be a paraphrase of another, etc. We see how the current proposal does this: two lexemes are synonymous iff they are mapped on isomorphic machines. Since finer distinctions largely rest in the *eidopoios diaphora* that we blatantly ignore, there are many synonyms: for example we define both *poodle* and *greyhound* as `dog`.

Second, we wanted the theory of lexical semantics to connect to a theory of the meaning of larger (non-lexicalized) constructions including, but not necessarily limited to, sentential syntax and semantics. The theory proposed here meets this criterion maximally, since it uses the exact same mechanism to describe meaning starting from the smallest morpheme to the largest construction (but not beyond, as communicative dynamics is left untreated).

Third, we wanted the theory to provide a means of linking up meanings across languages, serving as a translation pivot. While making good on this promise is obviously beyond the scope of this paper, it is clear that in the theory proposed here such a task must begin with aligning the primitives **D** developed for one language with those developed for another, a task we find quite doable at least as far as the major branches of IE (Romance, Slavic, and Germanic) are concerned.

Finally, we said that the theory should be coupled to some theory of inference that enables, at the very least, common sense reasoning about objects, people, and natural phenomena. We don't claim to have a full solution, but we conclude this paper with some preliminary remarks on the main issues. The complexities of the logic surrounding lexemes are not exactly at the same points where we find complexities in mathematical logic. In particular *truth,* which is treated as a primitive notion in mathematical logic, will be treated as a derived concept here, paraphrased as 'internal model corresponds in essence to external state of affairs'. This is almost the standard correspondence theory of truth, but the qualification 'in essence' takes away much of the deductive power of the standard theory.

The mode of inferencing supported here is not sound. For example, consider the following rule: *if A' is part of A and B' is the same part of B and A is bigger than B, then A' is bigger than B'.* Let's call this the Rule of Proportional Size, RPS. A specific instance would be that children's feet are smaller than adults' feet since children are smaller than adults.

Note that the rule is only statistically true: we can well imagine e.g. a bigger building with smaller rooms. Note also that both the premises and the conclusion are defeasible: there may be some children who are bigger than some adults to begin with, and we don't expect the rule to hold for them (this is a (meta)rule of its own, what we will call Specific Application), and even if the premises are met the conclusion need not follow, the rule is not sound.

Nevertheless, we feel comfortable with these rules, because they work most of the time, and when they don't a specific failure mode can always be found e.g. we will claim that the small building with the larger rooms, or the large building with the smaller rooms, is somehow not fully proportional, or that there are more rooms in the big building, etc. Also, such rules *are* statistically true, and they often come from inverting or otherwise generalizing rules which are sound, e.g. the rule that if we build A from bigger parts A' then B is built from parts B', A will be bigger than B. (This follows from our general notion of `size` which includes additivity.)

Once we do away with the soundness requirement for inference rules, we are no longer restricted to the handful of rules which are actually sound. We permit our rule base to evolve: for example the very first version of RPS may just say that big things have big parts (so that children's legs also come out smaller than adults' arms, something that will trigger a lot of counterexamples and thus efforts at rule revision), the restriction on it being the same part may only come later.

Importantly, the old rule doesn't go away just because we have a better new rule. What happens is that the new rule gets priority in the domain it was devised for, but the old rule is still considered applicable elsewhere.

Acknowledgements

We thank Tibor Beke (UMass Lowell) for trenchant criticism of earlier versions.

References

1. Allen, B., Gardiner, D., Frantz, D.: Noun incorporation in Southern Tiwa. IJAL 50 (1984)
2. Allen, J., Ferguson, G.: Actions and events in interval temporal logic. Journal of logic and computation 4(5), 531–579 (1994)
3. Aronoff, M.: Word Formation in Generative Grammar. MIT Press, Cambridge (1976)
4. Aronoff, M.: Orthography and linguistic theory: The syntactic basis of masoretic Hebrew punctuation. Language 61(1), 28–72 (1985)
5. Blackburn, P., Bos, J.: Representation and Inference for Natural Language. In: A First Course in Computational Semantics, CSLI, Stanford (2005)
6. Boguraev, B.K., Briscoe, E.J.: Computational Lexicography for Natural Language Processing, Longman (1989)
7. Cawdrey, R.: A table alphabetical of hard usual English words (1604)
8. Dowty, D.: Word Meaning and Montague Grammar. Reidel, Dordrecht (1979)
9. Dowty, D.: Thematic proto-roles and argument selection. Language 67, 547–619 (1991)
10. Eco, U.: The Search for the Perfect Language. Blackwell, Oxford (1995)

11. Eilenberg, S.: Automata, Languages, and Machines, vol. A. Academic Press, London (1974)
12. Fillmore, C., Atkins, S.: Framenet and lexicographic relevance. In: Proceedings of the First International Conference on Language Resources and Evaluation, Granada, Spain (1998)
13. Fillmore, C., Kay, P.: Berkeley Construction Grammar (1997), `http://www.icsi.berkeley.edu/~kay/bcg/ConGram.html`
14. Fillmore, C., Atkins, B.: Starting where the dictionaries stop: The challenge of corpus lexicography. Computational approaches to the lexicon, 349–393 (1994)
15. Flickinger, D.P.: Lexical Rules in the Hierarchical Lexicon. PhD Thesis, Stanford University (1987)
16. Graham, A.: Two Chinese Philosophers, London (1958)
17. Gruber, J.: Lexical structures in syntax and semantics. North-Holland, Amsterdam (1976)
18. Hayes, P.: The naive physics manifesto. Expert Systems (1979)
19. Hobbs, J.: Deep lexical semantics. In: Gelbukh, A. (ed.) CICLing 2008. LNCS, vol. 4919, pp. 183–193. Springer, Heidelberg (2008)
20. Keenan, E., Comrie, B.: Noun phrase accessibility and universal grammar. Linguistic inquiry 8(1), 63–99 (1977)
21. Keller, H.: The story of my life. Dover, New York (1903)
22. Kiparsky, P.: From cyclic phonology to lexical phonology. In: van der Hulst, H., Smith, N. (eds.) The structure of phonological representations, vol. I, pp. 131–175. Foris, Dordrecht (1982)
23. Kiparsky, P.: On the Architecture of Pāṇini's grammar. ms., Stanford University (2002)
24. Kipper, K., Dang, H.T., Palmer, M.: Class based construction of a verb lexicon. In: AAAI-2000 Seventeenth National Conference on Artificial Intelligence, Austin, TX (2000)
25. Kirsner, R.: From meaning to message in two theories: Cognitive and Saussurean views of the Modern Dutch demonstratives. Conceptualizations and mental processing in language, 80–114 (1993)
26. Kornai, A.: Mathematical Linguistics. Springer, Heidelberg (2008)
27. Kornai, A.: The treatment of ordinary quantification in English proper. Hungarian Review of Philosophy 51 (2009) (to appear)
28. Lenat, D.B., Guha, R.: Building Large Knowledge-Based Systems. Addison-Wesley, Reading (1990)
29. Levin, B.: English Verb Classes and Alternations: A Preliminary Investigation. University of Chicago Press, Chicago (1993)
30. Lewis, D.: General semantics. Synthese 22(1), 18–67 (1970)
31. McCarthy, J.: An example for natural language understanding and the ai problems it raises. Formalizing Common Sense: Papers by John McCarthy. Ablex Publishing Corporation 355 (1976)
32. McCarthy, J.J.: Human-level AI is harder than it seemed in 1955 (2005), `http://www.formalstanford.edu/ jmc/slides/wrong/wrong-sli/wrong-slihtml`
33. Mitchell, T.M., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., Just, M.: Predicting human brain activity associated with the meanings of nouns. Science 320(5880), 1191–1195 (2008)
34. Montague, R.: Universal grammar. Theoria 36, 373–398 (1970)
35. Ogden, C.: Basic English: a general introduction with rules and grammar. K. Paul, Trench, Trubner (1944)

36. Pustejovsky, J.: The Generative Lexicon. MIT Press, Cambridge (1995)
37. Quillian, M.R.: Semantic memory. In: Minsky (ed.) Semantic information processing, pp. 227–270. MIT Press, Cambridge (1967)
38. Quillian, M.R.: Word concepts: A theory and simulation of some basic semantic capabilities. Behavioral Science 12, 410–430 (1968)
39. Ruhl, C.: On monosemy: a study in lingusitic semantics. State University of New York Press (1989)
40. Schank, R.C.: Conceptual dependency: A theory of natural language understanding. Cognitive Psychology 3(4), 552–631 (1972)
41. Tarski, A., Givant, S.: A formalization of set theory without variables. Amer. Mathematical Society (1987)
42. Trier, J.: Der Deutsche Wortschatz im Sinnbezirk des Verstandes. C. Winter (1931)
43. Turner, R.: Montague semantics, nominalisations and Scott's domains. Linguistics and Philosophy 6, 259–288 (1983)
44. Turner, R.: Three theories of nominalized predicates. Studia Logica 44(2), 165–186 (1985)
45. Wierzbicka, A.: Lexicography and conceptual analysis. Karoma, Ann Arbor (1985)
46. Williams, E.: On the notions *lexically related* and *head of a word*. Linguistic Inquiry 12, 245–274 (1981)