Rebecca Green  
College of Information Studies, University of Maryland, USA

**Conceptual Universals in Knowledge Organization and Representation**

**Abstract:** Within the overall conference theme—integration of knowledge across boundaries—an important subtheme is universality: Where universals of knowledge organization and representation exist, knowledge integration is more likely. Thus, knowledge of conceptual universals should inform efforts at knowledge integration. In this paper, natural language is used as a model for exploring conceptual universals, since the phenomenon of translating between languages validates, but also circumscribes, the existence of semantic and lexical universals. The paper explores a representative inventory of semantic and lexical universals that should be accounted for in knowledge organization and representation systems, especially those that aim to be comprehensive.

1. **Introduction**

One of the biggest challenges we face in knowledge organization and representation is heterogeneity in the expression and structure of conceptual content. We have developed a vast array of classificatory systems, for example, classification schemes, thesauri, ontologies, and lexical databases. Their number exacerbates a problem that arises naturally from their individuality, that is, (some degree of) incompatibility. Responding to this situation, we have also developed a number of tools that show correspondences across classificatory systems, some of the same type, some of different types. Examples include: conversion tables between the Library of Congress Classification and Dewey Decimal Classification schemes (Scott, 1999); correspondences within Library of Congress Subject Headings to equivalent classes within the Library of Congress Classification; cross-language equivalences built into multilingual thesauri; the mapping of dozens of medical vocabularies to the Metathesaurus® of the Unified Medical Language System® (http://www.nlm.nih.gov/research/umls/); and mappings between versions of the WordNet lexical database (e.g., http://www.cs.ucf.edu/~bmartin/wnmap/). 

But even where mappings across classificatory systems exist, we are not guaranteed that the end result will manifest unified semantic content and structure.

Not surprisingly, the situation that occurs with classificatory languages—on the one hand, some correspondences cut across systems, but, on the other hand, individual systems exhibit distinctive characteristics—mirrors the situation that occurs with natural languages: We can usually translate satisfactorily between them, but recognize that we have often not said exactly the same thing in doing so. On the one hand, we perceive benefit in the many natural languages and classificatory systems that exist; if but a single natural language or a single system of classification existed, we would lose the richness associated with variety; moreover, since specific language systems tend to emphasize different arenas of experience, we would probably lose the ability to reflect in depth on some aspects of our world. Multiplicity of language and classificatory systems enriches our ability to express
conceptual content in both horizontal and vertical planes. On the other hand, if our natural languages and classificatory systems were distinctive to the point of complete dissimilarity, the existence of different natural languages and classificatory systems would stand as barriers to communication. The intellectual wealth associated with the multiplicity of language and classificatory systems would then be inaccessible to us. As a result, the situation we find with both natural languages and classificatory systems—in which many distinctive systems exist, which correspond in part, but probably only in part, to each other—is probably desirable and perhaps even optimal.

The question remains whether, or better, to what degree, it is possible to create an interlingua or switching language to which natural languages or classificatory systems could be fully mapped. Indeed, there have been numerous attempts to construct precisely such languages. It is not part of my agenda to analyze any specific such language system, but rather to explore desiderata for such tools. More specifically, my intent is to discuss some of the more important and central conceptual universals of knowledge organization and representation, with an eye to enumerating a criterial set of universal phenomena that any interlingua or switching language would have to accommodate.

2. Language Universals as a Model for Knowledge Organization and Representation

Because of the close relationship between natural languages, on the one hand, and knowledge organization and knowledge representation systems, on the other hand, a brief survey of background issues relating to natural language universals promises to illuminate our understanding of conceptual universals for knowledge organization and knowledge representation. The survey will first touch on general issues and then turn to the more specific context of translation.

2.1 Language Universals: General Background

A major landmark in the study of language universals was the 1961 Conference on Language Universals. A memorandum prepared for the conference begins, “Underlying the endless and fascinating idiosyncrasies of the world’s languages there are uniformities of universal scope. Amid infinite diversity, all languages are, as it were, cut from the same pattern” (Greenberg, Osgood, & Jenkins, 1966, xv). These uniformities and common patterns, like all collections of any size or complexity, admit of different ways of being classified. One way to slice the classification pie is to consider in what stratum of language a universal operates, e.g., phonology, grammar, lexicon, and so on. Another important dimension of language universals is what Greenberg et al. call their “logical structure” (pp. xix-xxi): Do they apply to all languages (unrestricted universals) or only to some (typically, many or most) languages (statistical universals)? Do they involve a relationship between two features, such that if a language has characteristic a, it will also have characteristic b (implications), or do they apply in all circumstances ([true] universals)?

All effective knowledge organization and representation schemes impose equivalence control between various expressions of (more or less) the same conceptual content. This process reflects a concern with cognitive content on two levels. First, there is the semantic level, which addresses what counts as a conceptual unit and how conceptual units are interrelated. Second, there is the
lexical level, which addresses how conceptual content is expressed in words and phrases. Consequently, it will be language universals operating in strata related to meaning—semantic and lexical universals—that are most relevant to our inquiry. Happily for our enterprise, Immler (1991, 38) suggests that “hundreds of valid linguistic universals [could be named] as soon as we turn to the domain of semantic universals.” A significant research program has also been pursued by Wierzbicka and others (Wierzbicka, 1992; Goddard & Wierzbicka, 1994) to identify lexical primitives—concepts that cannot be defined in terms of other concepts and that appear to be expressible in all human languages; such primitives constitute lexical universals. Figure 1 summarizes the set of lexical items (as expressed in English) found to appear cross-linguistically and assumed to be universal.

<table>
<thead>
<tr>
<th>Substantives</th>
<th>I, you, someone, something, people</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental predicates</td>
<td>think, say, know, feel, want</td>
</tr>
<tr>
<td>Determiners/quantifiers</td>
<td>this, the same, other, one, two, many, all</td>
</tr>
<tr>
<td>Actions/events</td>
<td>do, happen</td>
</tr>
<tr>
<td>Meta-predicates</td>
<td>no, if, can, like, because, very</td>
</tr>
<tr>
<td>Time/place</td>
<td>when, where, after, before, under, above</td>
</tr>
<tr>
<td>Partonomy/taxonomy</td>
<td>have parts, kind of</td>
</tr>
<tr>
<td>Evaluators/descriptors</td>
<td>good, bad, big, small</td>
</tr>
</tbody>
</table>

Figure 1. Wierzbicka’s proposed lexical and semantic primitives (Goddard, 1994, 22)

Turning to the logical structure of universals, one might initially assume that only unrestricted lexical and semantic universals—non-implicational assertions holding across all time and all languages—should matter in our pursuit of conceptual universals for knowledge organization and representation. There are several reasons to want to relax this assumption. First, the notion of unrestricted universals may be taken to presuppose a not-altogether-realistic inductive approach for postulating a universal, in which a pattern can only be claimed to be a universal if it is known to exist in all languages. But do we have full knowledge of all extant languages? What of languages that have died before being discovered by any linguist? Can we hope to predict how languages might change in the future? Second, if we somehow did have perfect knowledge of all unrestricted universals, we should presumably still want to account for the merely statistical universals—generalizations applying to many, but not all, languages—which are likely to predominate; Ullmann (1966,p. 220) claims that “most semantic universals are likely to be of the statistical variety.” Ullmann (p. 220) also suggests that “parallel developments” (e.g., certain metaphorical extensions that occur in some number of unrelated languages) warrant attention.

Immler (1991,p. 39) presents an alternative viewpoint, arguing that unrestricted semantic universals need not be based on inductive generalization: “As soon as we know the reason for the necessary presence of a feature in human language, it follows automatically that this feature must be present in all languages of the world; we then know that it must be a linguistic universal.” What would justify characterizing some conceptual content as a necessary feature of all human language? Immler takes up a variety of meanings purported to exist in all human
languages—for example, the distinctions between light and dark and between alive and dead—and then suggests “no human society could afford not to be able to express these ideas in its language” (pp. 39-40). He further argues that “cognitive contents can be universal across all cultures and all languages—they will be universal to the extent that the real world around us[—]which is the same for humans all over the world[—]is the object of cognition” (p. 49). Immler thus bases his claim for necessary semantic universals in the shared experience of human beings in the real world. Other linguists ground the universality of specific semantic components of language in human physiology, going so far as to characterize such components as “biologically given” and “innate” (Fillmore, 1971, p. 372 and Bierwisch, 1967, p. 3-4, respectively; cited in Goddard, 1994, p. 16). Another linguist maintains that basic semantic types, such as are reflected in the distinction between noun and verb, are “determined by the language-independent structure of the world (i.e., ontologically)” (Lyons, 1989, 161; cited in Goddard, 1994, p. 17).

To summarize: Investigation of semantic and lexical universals in natural languages can be expected to inform our exploration of conceptual universals for knowledge organization and representation schemes. The claim of “hundreds” of valid semantic universals and the presence of active research in the area of lexical universals lead us to expect that we may likewise posit a non-trivial set of conceptual universals for knowledge organization and representation. These universals appear to be grounded in human motivations, functions, activities, etc., stemming from our interaction with external, real world.

2.2 Language Universals: Translation

The extent to which we can satisfactorily translate between human languages is, it would seem, a good measure of the degree to which natural language incorporates semantic and lexical universals. If, on the one hand, whatever we can say in one language can be translated into (all) other languages, we might take that as prima facie evidence of an extensive array of unrestricted semantic universals; if the translation is also felicitous, we might also take that as prima facie evidence of the existence of an extensive array of lexical universals. If, on the other hand, all translations were felt to be approximations at best, there would be little hope for establishing either real semantic or real lexical universals.

Not surprisingly, there are proponents on both sides of the issue: “There are essentially two points of view from which translatability has been traditionally approached: the universalist one and the monadist one. Supporters of the former approach claim that the existence of linguistic universals ensure translatability. Those who endorse the latter approach maintain that each linguistic community interprets reality in its own particular way and this jeopardizes translatability” (Pedro, 1999). To the extent that a consensus has emerged between these positions, both extremes are rejected in favor of a compromise position in which “absolute untranslatability” is deemed not to hold and “perfect translation” is deemed “unattainable.”

Reflecting this intermediate position, Wierzbicka (1992, p. 7) notes that translation is possible, pointing to the large number of translations (“more than one thousand,” thus representing roughly one of every five or six extant languages) of (parts of) the Gospels. She also mentions, by way of contrast, the stock criticism leveled at translators: “Traddutore traditore” (“the translator is a betrayer”), citing the view that “every language provides its own set of lexicalised concepts, . . . [and] suggests its own categorization and its own interpretation of the world—
consequently, every language is indeed a different ‘guide to reality’ (Sapir 1949,162)” (Wierzbicka, 1992, p. 20). Weaving her way between the two extremes, she argues for a somewhat limited set of lexical universals (as seen in fig. 1), which presuppose underlying semantic universals: “There are good reasons to believe that every language has words available for the basic human concepts, and that everything that can be expressed at all can be expressed by combining those basic concepts in the right way. In this sense—but only in this sense—anything that can be said in one language can be translated, without a change of meaning, into other languages” (p. 20).

Philosophical concerns are largely absent from work on machine translation. Here the assumption of basic translatability is accepted even in the face of explicit recognition of the lexical, syntactic, semantic, and pragmatic differences between languages that make the task difficult. The most established approaches to machine translation incorporate assumptions relating to lexical and semantic universals, respectively. In its most naive version, the direct method involves word-for-word translations; more sophisticated versions of the direct method include various pre- and post-processing stages, but at some point, they, too, involve identifying a lexical unit in the target language that is considered equivalent to a word or phrase found in the source language. The direct method is thus based on the general assumption that for most pairs of languages, it will be possible to identify lexical units in the two languages that are conceptually equivalent. Generalized over a large number of languages, this becomes an assumption of the existence of lexical universals across many languages. Rather than identify translational equivalents between two specific languages, the indirect method performs translation by capturing the semantic content of the source language in an interlingual representation and then generating the target language translation from that semantic representation, a more efficient approach when many languages are involved. The indirect, interlingual method is thus based on the general assumption that the conceptual content expressible in any language can be captured in some one language, although it may be artificial. This intermediate language is in essence the language of semantic universals.

In summary, although translation theorists have come down on both sides of the fence on the question whether anything expressible in one language can be expressed in every other, or even any other, language, empirical translation results suggest a more moderate stance. Lexical and semantic universals may be expected to exist, but may be somewhat limited in scope or generality. The degree to which lexical and semantic universals are constrained in human translation also limit the effectiveness of machine translation, depending on the method used.

3. Semantic and Lexical Universals for Knowledge Organization and Representation

In enumerating conceptual universals that must be accommodated by any comprehensive knowledge organization or knowledge representation system, we may productively look to the semantic and lexical universals found in natural languages. Where empirical data are available, we will adopt an inductive method, although without any serious attempt to validate a universal in all human languages; generally the validation of a pattern in some number of unrelated languages is
assumed to suffice. But we will also adopt the methodology of accepting as
universals those phenomena that are grounded in shared human experience of the
real world, which “no human society could afford not to be able to express.”

3.1 Predicate-Argument Structure

Sasse (1991, 93) suggests that conceptual universals are more likely to exist
as “universal semantic principles” than as any kind of what is here termed a lexical
universal. One such principle is that of predication, which Sasse introduces in the
context of the core functionality of sentences, which is to express propositions: “A
proposition is normally based on the conception of an event or state, in which
certain individuals may or may not be involved. . . . As far as its content is
concerned, we would thus have two basic constituent parts of the sentence, the
event or state . . . and the individuals . . . which take part in it. Between the [two] . .
. there is a logical relation, which could be called the propositional relation,
presumably identical with the concept of predication in predicate logic” (p. 77). We
note that predication in first-order logic stands on the twin shoulders of predicates
and the terms they take as arguments. Similarly, the data modeling theory currently
in widest use recognizes two core types, entities and relationships, where
relationships are defined as associations between entity classes. Relationships can
readily be seen as predicates whose arguments are the entity classes they bring into
association. The interaction between relationships and entities is thus analogous to
the interaction between predicates and arguments. This yields three distinct types of
representational systems —natural language, predicate logic, and entity-relationship
modeling— all based on a similar bipartite conceptual structure, specifically, a
predicate-argument structure.

Immler (1991, p. 42) notes that there are languages where it is not possible to
analyze the expression of an event into predicates and arguments or where there is
no single linguistic unit that performs conjunction. However, sentences from these
languages can be translated into English where “the sentences as a whole are
perfect (or maximal) translational equivalences one of the other” even though there
is no one-to-one correspondence between the structural elements of the sentences. If
predicate-argument structure is universal, it occurs on the conceptual level and is
not necessarily reflected as such on the syntactic level.

However, a closely related syntactic universal has been proposed, namely,
that all languages have nouns and verbs (Dixon, 1977, p. 28; cited in Thompson,
1988, p. 168). Verbs are often used to express predicates, while nouns often occur
as their arguments. Not surprisingly, among Wierzbicka’s proposed inventory of
lexical and semantic universals (see fig. 1) are both substantives/nominals (I, you,
someone, something, people) and predicates, including those expressing
actions/events (do, happen) and those expressing mental predication (think, say,
know, feel, want) (Goddard, 1994, p. 22). Thus, even if the universality of
predicate-argument structure cannot be supported strictly on syntactic grounds, still
it is so central to our conceptual systems that in many, if not in most, cases the
verb/noun distinction will reflect a corresponding predicate/argument distinction.
Indeed, not only might we assume that all languages have some means for
expressing predicate-argument structure, but we might also assume that almost all
language use attempts to express predicate-argument structure that occurs at the
conceptual level. Accordingly, our investigation of further universals will be guided
by the predicate-argument distinction.
3.2 Characteristics of Arguments: Individuation, Classification, and Countability

We tend to think of that which exists in the real world as entities or things, typically corresponding to the argument side of the predicate-argument distinction. To the extent that the real/physical world exists independently of or logically prior to human cognition and to the extent that such real phenomena exist throughout the world, they are likely to occur as lexical universals—not necessarily in the narrow sense of being lexical primitives, but in the broader sense of being expressible (and probably lexicalized) in most/all languages. Immelr (1991, p. 39, quoting Immelr, 1974, p. 41) proposes that “in every language of the world there are names, designations for the following objects . . .: ‘soil, [many animals], [many plants], [parts of the body], . . . air, water, rain/snow, wind, sun.’” In like manner, Antal (1963, p. 85; quoted in Immelr, 1991, p. 49) suggests: “Although the world of meanings is different and theoretically incomparable and incommensurable from one nation to another, the denotata amongst which people live are, in fact, the same for all humanity or, at least, for a great number of people.”

The previous claim that “to the extent that . . . real phenomena exist throughout the world, they are likely to occur as lexical universals” is on some level naive, since the lexicalization process is driven in large part by human cognition/experience. Thus, the omnipresence of sun, moon, stars, and sky for all humans does not lead necessarily to a shared conceptualization of these phenomena. One can imagine the (at least logical) possibility of seeing them as fewer than, or more than, four types of things. And yet one can also imagine that most cultures will see them as exactly four types of things.

At issue here is, as a first step, the universality of what constitutes a single entity and, as a second step, the universality of which entities are seen as the same type of thing, i.e., belong to the same class. Full universality at the second stage is severely constrained not only by the partial nature of the distribution of things throughout the world (for example, some climates never experience frost or snow; flora and fauna may be confined to certain regions of the world), but also (and as importantly) by cultural and individual differences in knowledge, values, resources, etc. Still, we can expect where phenomena exhibit salient and discrete characteristics (of, for example, shape and color) and are of a size to make them easily observable by humans, universality of perception of what is a single entity unit is the more likely. As there are multiple such entity units which differ from each other in small ways, but which are distinct from other entity units (i.e., differ from them in significant ways), universality of perception of what belongs in the same class is also the more likely. Some things and groups of things in our world fit this description, but many others do not. Hence we can expect universality in these arenas to exist, but only in limited areas. Universality of class membership is of necessity constrained by universality in unit perception.

Although the universality of unit perception and the universality of class membership operate in only restricted arenas, the underlying phenomena—individuation and classification—are universal; and dependent on both individuation and classification we have another universal, that of countability.

Individuation is the principle of picking out specific regions within our world(s) and endowing them with independent ontological status, the principle of particularization or identification. Built into the principle of individuation is a basic distinction between what is the same thing and what is a different thing. Although
there are many situations in which we need to deal on the level of particulars—one should be careful, for example, not to drive away in any vehicle other than the specific one to which one holds the title, no matter how similar they may be—there are many other situations (probably far many more) in which individuation is neither necessary nor desirable. Any appropriately valued token or set of tokens may be used in a telephone booth, for example. It is not simply for purposes of cognitive economy that we so often deal at the level of classes rather than at the level of individual entities; much of our reasoning is done at the class level (Sokal, 1974). Thus we would expect the distinction between classes of entities and individual entities, that is, the type/token distinction, to be universal.

Once we recognize both individual entities and classes comprised of multiple entities of the same kind, we come quickly to the notion of quantifying the membership of the class: How many individual entities are perceived as being of a kind? This quantification is effected through counting individual entities. A related process of quantification can be applied to entity regions that do not undergo individuation; these are perceived in terms of mass quantity, not count quantity. In this case units of measure (e.g., liters, cups, handfuls) substitute for individual entities as countable items, individuation of some sort being a prerequisite to counting (Link, 1991, p. 133). In either case, whether we perceive the entity being measured in terms of mass or count quantification, countability produces a universal set of integers. Another extension of countability gives us the difference between ‘some’ and ‘all’, the distinction between existential and universal quantification.

A personal application of individuation generates a sense of self that is distinct from everything else in all ‘normal/healthy’ persons (this sense of self would generally be considered criterial for normalcy and health). Consequently, all languages can be expected to have some notion equivalent to first person singular. The application of classification to each I results in first person plural, we, which is likely to admit of different memberships in different contexts. Meanwhile, all persons not part of us are, collectively, them, third person plural, or, individually, he/she/it (the gender breakdown for third person not being established as a universal by anything examined thus far; similarly, the recognition of second person arises from the phenomenon of communication and is thus less philosophically central than first and third persons).

In summary, for substantives we accept the full universality of several basic principles—individuation, classification, and countability. Furthermore, we assume the universality of some limited number of real world phenomena.

3.3 Characteristics of Predicates: States, Events, and Roles

On the predicate side of the predicate-argument distinction, we have a potentially large set of conceptual relationship types, typically broken down into states and events, with states as static relationships and events as dynamic relationships (although the breakdown is not nearly so clean as suggested: where, for example, do such cognitive predicates as believing, remembering, and recognizing belong?).

We discussed previously the real world grounding of some entities that are likely to be universally recognized, for example, the sun. Such universals are based on the omnipresence of certain substantives in the experience of essentially all persons. A parallel argument can be made for the universal recognition of various physical states that exist in nature and that are likely to impinge on the life of all
persons, for example, light vs. dark, wet vs. dry, warm vs. cold, alive vs. dead, big/large/wide/deep vs. little/short/narrow/shallow, hard vs. soft, heavy vs. light, smooth vs. rough, young vs. old, sick vs. healthy, pain vs. pleasure (Immler, 1991, p. 39 [also citing Inunler, 1974, p. 41]; Thompson, 1988, p. 168); the latter few states are more likely to be known from personal experience, although many persons would also recognize them from animal life around them.

Other stative universals seem intrinsic to human physiological experience. These include such states as hunger, thirst, weariness, and a specific set of emotions (joy, sorrow, fear, disgust, anger, surprise) whose facial expressions are recognized cross-culturally (Ekman, 1993). Similarly, although color boundaries are language-specific, the perception of focal colors is apparently universal (Kay, 1999).

Additional states of a universal nature arise from the human social experience. Such might include basic kinship relationships (at least the general notion of kinship relationships) and evaluative properties (good vs. bad).

Dynamic universals would probably similarly arise from real world phenomena (e.g., weather events, manners of locomotion—running, flying, swimming—observed in animals), from human physiology (e.g., eating, drinking, sleeping; vision, touch, taste), and the basic social, cognitive, and psychological nature of humans (e.g., verbal communication, birth, belief/thought/memory, desire).

Predicates are often characterized in terms of the semantic roles of their arguments, where a semantic role is a functional semantic type at the most abstract level. Attempts have been made to enumerate a universal set of semantic roles, but the endeavor is probably ill-fated. Davis (1994, p. 161, 163) suggests that recognizing their “constant presence in language” is a more productive approach to semantic roles than trying to develop a universal inventory.

However, as we have already repeatedly seen, not being able to corroborate that a comprehensive set of universals of some sort exists is not tantamount to denying that any universals of that sort exist. For example, Comrie (1989, p. 59) notes with respect to the roles of agent, force, instrument, experiencer, and patient that they are “not so much a set of discrete semantic relations, but rather a continuum . . . of control.” These roles hold central positions in those semantic role inventories that have been proposed. We may therefore assume that, even though they may not constitute discrete cross-linguistic semantic roles, control is probably a central notion in all languages. If so, we may also assume the universal centrality of animacy, which correlates closely with control.

To the extent that roles can be identified with general universal predicates—for example, event or state, at the most general level, or communication or locomotion, at a somewhat more specific level—we may likewise find that those semantic roles are universal. Here we take the example of events, which take place in time and space. Given the linguistic universality of events, we would expect to find time and space as semantic roles in all languages. In a cross-cultural study of temporal collocations and metaphors in English, Mandarin, Hindi, and Sesotho, Alverson (1994) goes further, claiming that time is experienced and expressed in similar ways universally: “as a partible entity, . . . as a causal agent with effects, . . . as a medium in motion, . . . as a linear and as a circular course, . . . and . . . as its method of ascertainment” (p. 104). Likewise, although spatial expressions in different languages display syntactic variation, there are underlying patterns that cut across them: objects (trajectors) are located in relation to other objects (landmarks),
in terms of either the coincidence or separation of their locations, construed in 1-, 2-, or 3-dimensions (Hawkins, 1993).

In summary, for predicates we accept the full universality of some specific states and events—probably numbering in the dozens—that occur in the world and/or that arise from common human experience. Further, we recognize the universality of semantic roles; although it may be impossible to enumerate a well-defined and restricted set of roles, still some notions underlying semantic roles appear to be universal: control, as a universal parameter of roles, on the one hand, and time and space, as universal roles, on the other hand.

3.4 Logical Operators

So far, we have been operating in the context of a bipartite conceptual structure based on predicates and their arguments. Weinreich (1966, p. 148) takes this line of reasoning further when he claims, “It will be assumed that it is possible to describe all discourse as either (a) having the semiotic form ‘O(x)’, or (b) deviating from it in specified ways. In this formulation ‘x’ stands for an argument—“something talked about”; ‘O’ for a predicate—“something said about x”; and ‘O’ is a covering label for any of a number of operations.” Weinreich therefore proposes the universal existence of a set of operators, which supplement our fundamental categories of predicates and arguments. Some of Weinreich’s operations are strictly linguistic in nature and of no particular consequence to us here. Other notions that Weinreich treats in terms of operators (e.g., person, time, place, same vs. different, count vs. mass), we have handled elsewhere. There remain the propositional operations of negation, disjunction, conjunction, implication, and equivalence. Weinreich acknowledges that “no language represents such operations with the maximum economy” (p. 158), meaning that a one-to-one correspondence between linguistic expressions and logical operations is often lacking. Indeed some amount of circumlocution may be required to express such concepts; moreover, the means available for expressing them may include as well the expression of other semantic components. That it is logically possible to define all these operations in terms of negation and conjunction further obscures the picture. Nevertheless, we can propose reasons why most of these logical operations would be universal. Negation is implied by antonymy, which is considered a universal (see section 3.5). According to Weinreich, linking, an effect of conjunction, is one of two semiotic processes (along with nesting) by which linguistic signs can be “grammatically combined to form discourse” (pp. 163-164). If, as he claims, “in all languages a combination of signs takes the form of either linking or nesting, and all languages use both patterns in kernel sentences” (p. 167), then conjunction, too, must be universal. Disjunction (with its intimate relationship to choice) and implication (with its close connection to causation) probably are also universal.

3.5 Lexical Relationships

The standard paradigmatic/lexicosemantic relationships which must be handled in thesauri—taxonomy, partonomy, synonymy, antonymy—are likely to be found in all languages. This generalization implies that all natural lexicons are structured (at least eventually so) and that human-constructed languages are simply mirroring relationships that occur in natural language.

Evidence from empirical studies of many languages supports the universality of such hierarchical relationships as the subsumption/taxonomy/IS-A relationship and the part-whole relationship (Goddard & Wierzbicka, 1994). In a similar vein Ullmann (1966, p. 231) refers to cross-linguistic evidence for the universality of
synonymy wherever there is, as there is always likely to be, specialization: “It has often been found that subjects prominent in the interests and activities of a community tend to attract a large number of synonyms.” Similarly, the presence of antonyms cross-linguistically argues for the admission of antonymy as a language universal (Lehrer & Lehrer, 1982; Raybeck & Herrmann, 1996).

Ullmann argues for the universality of polysemy on the basis of both cognitive economy and the richness of language: “Polysemy is in all probability a semantic universal inherent in the fundamental structure of language. The alternative to it is quite unthinkable: it would mean that we would have to store in our brains a tremendous stock of words, with separate names for any possible subject we might wish to talk about; it would also mean that there would be no metaphors and that language would thus be robbed of much of its expressiveness and flexibility” (p. 232). But, according to Greenberg et al. (1966, p. xxii), “All languages have some metaphorically transferred meanings.”

4. Summary and Conclusion

What are the implications of this survey of semantic and lexical universals for the existence of conceptual universals in knowledge organization and knowledge representation systems and (2) for knowledge integration? First, we are led to reject two extremes. On the one hand, we cannot reasonably hope to build a super-system (whether it be a classification scheme, a thesaurus, an ontology) that faithfully integrates all extant instances of that type of scheme. Individual schemes may be unique in the set of substantives and predicates they recognize as well as how they relate conceptual units to each other. On the other hand, we also cannot responsibly throw up our hands, supposing we cannot do anything. There appear to be (dozens of) conceptual units (e.g., sun, bad, think) that are more or less universally apprehended. There are also some number of conceptual abstractions (e.g., predicate-argument structure, countability, animacy, hierarchy) that are used universally in structuring our conceptual systems. Any two knowledge-based schemes are likely to be alike in fundamental ways.

How can we build upon that which knowledge organization or knowledge representation schemes share with each other? Obviously, where specific conceptual units are shared, we should construct cross-system mappings; another paper at this conference (Green, Bean, & Hudon, this volume) reports preliminary data suggesting that this sharing is more likely to occur at the basic level than at hierarchical levels above or below the basic level. It may be that the best we can do in integrating two knowledge-based schemes is to construct such crosswalks as are truly appropriate and then rely on the individual structures of the schemes being “integrated.” (Indeed, this is the approach of several major efforts of this type, e.g., EuroWordNet and the Metathesaurus® of the Unified Medical Language System ®.)

As for the shared conceptual abstractions, they may be best used as parameters for analyzing any two systems that are subject to integration efforts. For example, what differences exist between the set of substantives treated as count nouns vs. those treated as mass nouns? Where do the semantic roles implied by the two systems fall on a control continuum? There is no magic key to knowledge
integration. However, knowledge of these conceptual universals can at least facilitate and illuminate the very considerable work that knowledge integration will always require.

Whether we can do more will be subject to continued exploration, but this much we can do: (1) Recognize true conceptual equivalences across systems; these are most likely to be found in those areas of knowledge-based systems that address real world phenomena at the basic level. (2) Retain the structural integrity of individual systems. (3) Use universal conceptual abstractions to analyze these systems, so as to further our understanding of the nature of their differences.

Notes
1. Some of the more recent approaches taken to machine translation (example-based translation, statistical translation) do not engage in isolating the meaning of a text in the source language before generating a meaning-equivalent expression in the target language; instead they depend on the pre-existence and availability of a large parallel corpus, in which extensive amounts of text have already been translated and in which corresponding units of source and target texts are aligned with each other. Since no meaningful semantic representation is established, these translation models do not further our understanding of semantic universals.

References


