Toward a Classification of Relationships*†

Rick Szostak

Department of Economics, University of Alberta, Tory Building 9-18, Edmonton, Alberta, T6G 2H4, Canada <Rick.Szostak@ualberta.ca>

Rick Szostak is professor of Economics at the University of Alberta. The focus of his research for the last fifteen years has been to facilitate interdisciplinary research and teaching. Knowledge organization has become the dominant element in that research agenda, and he has now authored two books and several articles that develop classifications of things studied, theories, and methods applied, types of data, ethical perspectives, research practices, and now relationships among things. He has contributed to the Integrative Levels Classification and is developing the Basic Concepts Classification. He has argued in several places for the value of a classification grounded not in disciplines, but in the things we study and the relationships among these. He is working with others on a book about interdisciplinarity and knowledge organization.

Szostak, Rick. **Toward a Classification of Relationships.** *Knowledge Organization. 39(2)*, 83-94. 29 references.

ABSTRACT: Several attempts have been made to develop a classification of relationships, but none of these have been widely accepted or applied within information science. It would seem that information scientists, while appreciating the potential value of a classification of relationships, have found all previous classifications to be too complicated in application relative to the benefits they provide. This paper begins by reviewing previous attempts and drawing lessons from these. It then surveys a range of sources within and beyond the field of knowledge organization that can together provide the basis for the development of a novel classification of relationships. One critical insight is that relationships governing causation/influence should be accorded priority.

Received 2 August 2011; Revised 30 September 2011; Accepted 31 October 2011

* Appendix: The Appendix can be found at: http://www.economics.ualberta.ca/en/FacultyandStaff/SzostakRick.aspx

† The author thanks Claudio Gnoli and Tom Dousa for helpful advice. The author also thanks Barbara Kwasnik, Vivien Petras, and Leila Rotschy for respectively advising him to consult Khoo, SUMO, and NSM. The editor and three anonymous referees each provided many useful suggestions.

1.0 Introduction

Information scientists have long recognized the advantages of classifying the relationships that exist among things. Several complex schemes for doing so have been proposed (see Perrault 1994 for a summary), but none of these have been utilized in any major classificatory scheme. Classificationists have apparently decided that the proposed schemes do not deliver enough classificatory benefit to justify the cost of mastery by classificationist, classifier, and user. This paper explores whether it is possible to develop a scheme that delivers more and/or is less complex.

The starting point for this paper is the simple observation that most scholarly research—and likely most general non-fiction—studies how one or more things affect one or more other things. Thus, by far the most important relationships that must be investigated by the classificationist are those that involve some sort of causation or influence. Any proposed scheme for classifying relationships that does not devote the bulk of its attention to causal relations will of necessity fail to maximize the value versus cost ratio. Note here that the word 'causal' is used in the broadest sense to refer to any instance where (it is alleged that) one thing exerts some influence on another; the

word 'causal' in no way implies that this influence need be large and certainly not that it is the only influence on the thing being affected.

This last point deserves emphasis. Even scholars who disdain words such as 'cause' or even 'influence' often speak of how one thing affects another (for example, how a work of art moves an audience). The classificationist needs to capture all of these different types of cause/influence. Likewise we need to embrace—and perhaps distinguish—different types of causation/influence identified by philosophers: individual instances (child kicks ball), causal laws (the laws of thermodynamics), and causal possibilities (aspirins can reduce headaches).

Philosophers also debate the grounds on which we make causal statements. When we see a child kick a ball, we infer from the movement of the child's leg and the subsequent movement of the ball that the child caused the ball to move. But arguably we have no solid basis for making this inference, but rather have chosen to organize our perceptions around the idea of causation. The information scientist need not enter this debate. Traditionally we have classified works with regard to what they are 'about' without feeling any need to pass judgment on subject matter: we can classify works on astrology without feeling that we thereby endorse astrology. Likewise we can classify the idea 'child kicks ball' without endorsing any particular philosophical attitude toward causation.

If the vast bulk of at least non-fiction works in the world are focused on one or more causal relationships, then it follows that the best way to guide users to the works they seek is to facilitate search by causal relationships. This will be true not just for scholarly searches but also popular searches (stopping dog from biting mailperson). While Boolean searches using keywords can be helpful here, subject searches in terms of all possible causal relationships would be far superior. And that would require that works were classified in terms of combinations of things and the relationships between them.

This work is motivated in particular by the needs of interdisciplinary scholars. These will often be interested in how things commonly studied in one discipline affect things commonly studied by another: how attitudes toward punctuality affect economic growth; what chemical interactions are implicated in biological processes, how works of art generate an emotional response, and so on. Such searches are notoriously difficult at present due to the fact that universal catalogues are organized by discipline and employ different terminology across disciplines (see Szostak and Gnoli 2009). The author's own academic transition from economic historian to working at the interface of interdisciplinary and information science scholarship owes much to the difficulty he faced in locating relevant works by sociologists or political scientists that would discuss how particular cultural attitudes or institutions might influence economic and technological behaviors in particular ways.

The importance of causal relationships, broadly defined, has often been stressed in the Knowledge Organization literature. The excellent survey by Bean, Green, and Myaeng (2002) speaks of three broad types of relationship: equivalence (the main focus of thesauri), hierarchical, and associative. They note that there is no agreement on types of associative relationship, but laudably focus their attention on cause-effect relationships. Zeng, Zuber, and Salaba (2010) provide what they believe is an exhaustive list of types of associative relationship that should be—but often are not—captured in subject authorities. One of their ten types is hierarchical (whole/ parts), and another two can generally be captured by the non-causal relator 'of' (object/ field of study and concept/properties). The rest are each a type of or component of a causal relationship: cause/effect proper, the action/ process that an agent undertakes (speedometer measures), the result of that action (cloth woven), the agent that is affected (student taught), counter-agent (pesticides control pests), raw material (wine is made from grapes), and properties of actions (communicates well).

We thus have our first guiding principle: Principle 1: "Focus on causal relationships (influences)." As we proceed through the paper, other principles will be identified that might allow the development of a classification of relationships whose benefits exceed its costs.

Though the utility of a classification of relationships may be greatest in concert with a universal classification of things such as in the Integrative Levels Classification (ISKO Italia n.d.) or the Basic Concepts Classification (Szostak 2011a, 2011c), increased clarity can be provided by adding a classification of relationships to any existing classification system by coding works also for type of relationship (see Szostak 2011c). Such a classification would also be a useful input into subject authorities and thesauri (again see Zeng, Zuber, and Salaba 2010 and Bean, Green, and Myaeng 2002). Also, formal ontologies and the semantic web arguably require a classification of causal relationships in order to capture the meaning of causal arguments. Topic maps (Melgar Estrada 2011) are yet

another approach that could potentially benefit from an exhaustive classification of causal relationships.

Of course, there are non-causal relationships as well. Yet to a considerable extent these have already been addressed out of necessity within existing classifications. The 'part of' or 'type of' relationship are the keys to any logical hierarchical organization of things. And virtually every classification has some notation for 'and,' 'of,' and 'for.' There is some value in fleshing out the list of non-causal relators, but it is still the case that classificationists need little more than the handful of these that are already in common usage.

Section 2 of the paper briefly reviews two modern detailed classifications of relationships: by Farradane and Perrault. It will be shown that both efforts focused too little on causal relationships. It is desirable to reflect briefly on the value of the non-causal relators they identify. With respect to both causal and non-causal relationships, lessons can be derived from these two attempts as to how best to proceed in developing a classification of relationships.

Section 3 clears some potential complications out of the way. Section 4 draws on a variety of sources to deduce further principles for the classification of relationships. These principles alone cannot generate the desired classification. Section 5 then performs a broad inductive survey of works both within and beyond Knowledge Organization that can provide a basis for such a classification. Section 6 concludes.

2.0 Reviewing Farradane and Perrault

Farradane (1967) develops a 3x3 table of 'generic relationships.' One axis in the table represents "Increasing clarity of perception," while the other captures "Increasing association (mental time)." Little space is devoted to explaining the rationale behind these axes, but brief descriptions are provided of the nine types of relationship generated:

Concurrence (juxtaposition). This mostly captures 'and' relationships, but also 'of' as in 'encyclopedia of chemistry.'

Self-activity. Examples include 'Man walking' and 'Bird migrating.' There are two useful insights here. Most importantly, we should appreciate that the verb-like terms used in causal relationships may also be applied internally to the acts or internal workings of one thing. Secondly, as Farradane appreciates, we would want to distinguish the two types of usage. However, whereas Farradane proposes a different notation for these two cases, the differences could in fact be signaled within linked notation simply by whether another term was included in the class title (sugar crystallizing versus factory crystallizing sugar). Critically, it is not necessary to develop a distinct classification of self-activity: one classification can potentially serve both self-activity and relationships.

Association. This captures a range of associations. The first examples Farradane provides here actually involve a causal relationship (Prison/disgrace; or Cathedral/beauty), or are examples of self-activity (Hydrolysis/acid or Cutting/knife). Later examples would best be captured by the use of 'of' or 'for' (centrifuge for sugar).

Equivalence. This relationship between synonyms is of course critical for thesauri, but all synonymous terms should occupy exactly the same place in a classification.

Dimensional. This denotes placement in space or time. Examples would include 'above' or 'temporary' or 'India.' The last should be captured within a comprehensive classification of places and time periods. The former examples deserve to be clarified within a classification of spatial and temporal relators.

Appurtenance. As Farradane notes, these 'whole/part' relationships are best captured within a logical hierarchic classification of things. He also includes here certain properties (Syrup/density) which might usefully be indicated by the common relator 'of.' It remains to be seen if it is useful to distinguish different uses of the 'of' relator.

Distinctness. Farradane notes that it is only rarely necessary to denote that one thing is different from another. Though he does not suggest it, the phrase 'versus' might usefully capture distinctness.

Reaction. This means the action of any one thing or process on any other thing or process. Farradane takes great pains here to distinguish examples of reaction from examples of self-activity or appurtenance. But these distinctions do not seem to be important to the user: they will care about whether neutralization affects alkalinity of a substance, but not whether this is properly appurte-

nance or reaction. Again, then, we can proceed to develop one classification of causal relationships, and this set of relators can be used whenever a causal relationship of any type is at work.

Functional dependence (or causation). This is when A causes B or B arises out of A. An example is Bread/Wheat. Note that even in this example wheat is not the sole 'cause' of bread. Attempting to distinguish cases of sole causation from cases of partial causation is unlikely to provide much service to the user: they will generally be interested in works about a particular relationship, whatever the degree of relationship posited by the author. We can once again proceed to develop one classification of causal relators. We might, at most, wish to add some way of indicating the degree of influence alleged by the author.

In sum, Farradane's classification makes distinctions that are hard to comprehend and that will provide little advantage to the user. Yet our analysis suggests that Farradane's goal of increased clarity in the expression of relationships can be met by one classification of causal relationships supplemented by a set of non-causal relators including 'and,' 'of,' 'for,' and a set of locational and temporal indicators.

Perrault (1994) also devotes far greater attention to non-causal than to causal relationships (as does Stock 2010, and to a lesser extent Nutter 1989). Of his three main classes, the two most developed address primarily non-causal relationships. Perrault attempts much greater subdivision of his main classes than had Farradane. Unfortunately, he assumes that it is always best to subdivide into two or three subclasses.

His first main class, the 'subsumptive,' addresses primarily the type/kind or whole/part relationship. As noted above, a logical hierarchy of things in which subdivision only occurred in terms of type/kind or whole/part subdivisions would largely obviate the need for symbolic representation of this type of relationship. The distinctions stressed by Perrault (genus/species versus species/individual as type/kind; organism/organ versus composite/constituent as whole/part) would generally be clear in context. Notably, a couple of the examples Perrault provides here (principle/manifestation and matrix/particles) would seem to be better captured by causal relationships. Perrault also addresses subject/property relationships. These would generally be captured in most classifications through the use of 'of.' Perrault distinguishes 'property of' from 'possession of.' This distinction would likely always be clear in context. Again Perrault provides an example (Substance/accident) which might better be handled through a causal relationship. He closes with a suggestion that in some cases we would want to indicate 'with,' 'without,' or 'accompanies.'

His third class, the ordinal, receives the most attention. His first subclass, the conditional, addresses whether A affects B, or vice versa, or both. This is properly a matter of causation, and could be handled by a general indicator of causation, with further notation capturing direction of effect. His second subclass, 'state,' is subdivided into necessary, arbitrary, and contingent. As noted above, one might develop some indicator of the strength of a particular causal relationship, though it is not clear how useful that would be. The third subclass is 'attitude,' with subclasses favorable, indifferent, and unfavorable. Many classifications have some way of capturing 'from the perspective of.' It may or may not be useful to indicate what type of attitude is at work. Mazzocchi et al. (2007) speak of how insects are pests from the point of view of farmers, but this attitude will much of the time be captured by a causal relation such as 'insects harm crops.' Still, for works focusing on attitudes (at least) we would want some way of denoting perspective. Perrault's fourth subclass, 'energy' addresses whether A is capable of B. It is not at all clear that this class is useful. If we wish to distinguish works that argue for a particular relationship from those that suggest otherwise, we need only a notational device that signals 'not.' His fifth class, 'comparative' captures the more/less/equal distinction, generally and with respect to size, length, and duration. Such distinctions may at times be useful to the classificationist. His sixth class, 'locational,' addresses issues of time and place. In addition to the relations noted above by Farradane, he adds inside, outside, between, near, far, parallel, right, left, front, back, middle, during, toward, at, and away.

His second main class, the 'determinative,' primarily addresses causation. He has three main subclasses, 'active,' 'interactive,' and 'passive,' each of which in turn is divided into three sub-subclasses which capture basically positive, limited, and negative effects. As has been suggested above, we can save ourselves from much undue classificatory complication by proposing some notational indicator for 'does not cause' or for 'has negative effect.' We can then proceed to develop a classification of generally positive relationships, and capture neutral or negative effects through the appropriate clarification. We can focus here on just the positive sub-subclasses. Within the

'active' subclass, Perrault attempts to distinguish causing, originating, and influencing. As was noted above, this distinction is inadvisable. In the 'interactive' subclass, Perrault identifies association, imitation, and cooperation. These are all important relationships. In the 'passive' subclass, Perrault again lists influenced (and also 'affected' and 'derived from'). As argued above, there is no need to distinguish self-active causation from external causation. In sum, then, Perrault's scheme suggests a needless subdivision of the basic term 'causation/influence' while providing only a handful of useful suggestions regarding precise types of causation.

While a logical structure is vital to any classification, Farradane's 3x3 matrix and Perrault's subdivisions by twos and threes almost guarantee that some parts of the classification receive too much attention and others too little. In both cases, it also proves difficult to appreciate the divisions between classes and thus to readily place materials in the correct class.

Principle 2 follows: before any class is admitted to the classification, it will be asked whether the increased complication of the classification is justified in terms of value to classificationist, classifier, and/or user. A key subsidiary question is whether it is easy for classificationist, classifier, and user to apprehend the meaning of—and thus likely boundaries to—particular classes.

3.0 Potential Barriers to Classifying Causal Relationships

This paper began by suggesting that classifications of relationships had erred in under-emphasizing causal relationships. This section addresses some needless complications that might hinder the classification of causal relationships themselves.

3.1 Relationships and Facets

Vickery (2008) performs an important service in urging information scientists to distinguish 'facets' from 'relationships.' Most advocates of facet analysis treat relationships as one or more kinds of facets. Yet there is a clear logical distinction between aspects of one thing and relationships between two or more things. Ranganathan had appreciated this when he developed the idea of 'phase relations,' but there is still a tendency to treat these as just another kind of facet.

The starting premise of this paper—that most of the works (and ideas) we wish to classify involve some causal relationship—adds force to Vickery's advice. Efforts to squeeze this key element of document classification into something broader will likely lead to neglect. That being said, recall that verb-like terms can be used to describe both relationships between things and activities internal to one thing. If our classification of relationships proves useful also for capturing certain facets, we could hardly object to such an outcome.

3.2 Acting and Reacting

Both Farradane and Perrault emphasize a distinction between 'causing' and 'reacting.' I have stressed this precise distinction in a different context (Szostak 2003, 2004). Yet it is far from obvious that the classificationist need make this distinction. The statement "Cats react to stress by hiding" could be reworded as "Stressed cats hide." The act of hiding in the second is equivalent to the reaction of hiding in the first. Either way, reaction is clear in context. Likewise a general statement that "A encourages B to influence C" makes it clear that the impact of B on C is at least in part a reaction to the impact of A on B. Should it prove desirable in some contexts to distinguish reaction from action, this could be done by adding a notation for 'reaction' to the notation for any type of action. That is, since reactions involve the same set of verbs as actions, it is not necessary to develop separate classifications of these.

Similar arguments can be made with respect to distinctions between 'intended versus unintended,' 'enabling versus causing,' and 'focused versus peripheral.'

3.3 Unidirectional versus Reciprocal Influence

We will certainly want to signal the direction(s) of influence when coding particular documents. As with 'reacting' though, we can proceed to develop one classification of influences, and then employ some simple notational device to indicate direction(s) of influence.

3.4 Identifying Cases

Nor is it necessary to develop detailed classification of "cases." Grammarians seek to identify a set of cases or roles that a relationship might connect, such as agent, experience, instrument, object, or goal. Yet grammarians disagree on a precise list of such cases. (There is a loose connection here to the facets identified in different applications of facet theory.) Only if it were thought that the same relator took on quite

different meanings depending on the cases it connected, and such differences were not clear in context, would we need to develop separate case-specific classifications.

3.5 Necessary and Sufficient

As noted in section 2, we need at most here a notation that expresses the degree of causal influence posited within a certain work. We need not, in general, worry about this matter of degree in classifying. Note, though, that a word such as 'forced' means 'caused in a sufficient manner' while 'required' denotes 'caused in a necessary manner.' The classificationist may thus wish to indicate more specific instances of causation of a necessary or sufficient manner. Along with necessity and sufficiency, some indicator of probability of effect might be valuable: how often is a particular effect anticipated to occur? (Barriere 2002)

The same advice can be applied to issues of temporality. Does the causal relationship occur at one time or continuously? Does the relationship between cause and effect cycle?

A slightly different distinction does, however, merit attention. Barriere (2002) follows others in distinguishing causes that affect the very existence of a result from causes that merely affect some characteristic(s). In the first case, there are four possibilities: creation, destruction, maintenance, and prevention. All must be captured in our classification of relationships. In the second, change may occur in diverse ways, but the three most common are changes in size, duration, and length. Along dimensions such as these, the possibilities are increase, decrease, and maintain, while the more generic 'modify' can be used when the direction of change is unknown or does not exist. These sorts of causation obviously lend themselves to compounding.

3.6 Verbs are Numerous but perhaps Manageable

Though there are far fewer verbs than nouns in any language—and therefore it is reasonable to aspire to a much more compact classification of relations than of things—there are still well over two thousand verbs in the English language (Khoo 1995). As Khoo found, these can, for the most part, be grouped into classes of verbs with very similar meanings. This finding suggests that it is reasonable to seek a classification of relationships that contains far fewer than two thousand main entries. And we have argued that various types of compounding will reduce that total.

The experience of SYNTOL, an early effort to propose a classifier-assisted computer program for indexing and retrieval of works in any domain, is instructive. The developers of SYNTOL discovered very early that compound indexing solely in terms of things led to too many false drops in retrieval, presumably because users sought some particular relationship between two things. They found that even adding poorly specified relationship terms greatly facilitated retrieval; context generally clarified what otherwise could have been ambiguous (though note that the system was generally applied to domain-specific collections) (Gardin 1965, 53-5). As we shall see below, the (much better funded and developed) Unified Medical Language System also operates successfully with a one-page table of relationships.

4.0 Deductive Insights

Though Farradane's classification of relationships was problematic, he elsewhere in the same paper attempted to develop preliminary classifications of 'entities,' 'activities,' 'abstracts,' and 'properties,' and argued that these four together comprise the essence of what needs to be classified. The first and last of these classes comprise for the most part the noun-like things and adjective-like descriptors that account for the vast bulk of entries in existing classifications. The 'abstract' class is more troublesome. It includes the non-causal relators associated with time and space (above, below, before, after) that were discussed above. It also includes symbols (letters, numbers, words, sentences, etc.), which would generally be treated in the same way as 'entities.' Yet it also includes two subclasses of physical abstracts (rays, energy, heat, light) and behavioral abstracts (love, hate, pain) that—while they are in some ways 'entities'—are best conceived as relationships or at least self-activities: A heated B, C loves D, E feels pain.

Farradane's 'activity' class (supplemented by the physical and behavioral abstract subclasses) provides some idea of how to begin classifying causal relationships. He distinguishes physical activities (moving) from living activities (breathing), physical abstracts (increasing, using), and mental abstracts (counting, reasoning). Within each subclass, he identifies simple activities and complex activities. He also appreciates that complex activities may draw on simpler activities from more than one subclass: singing combines the physiological and mental.

Three of Farradane's main classes bear some similarity to three of Aristotle's four types of causation: Ar-

istotle's efficient cause encompassed physical actions, his final cause dealt with the purpose intended by an intentional agent, and his formal cause analyzed the internal structure of a causal agent and thus would include physiological behaviors. Aristotle's material cause (the material of an object causes its existence) might best be captured by the whole/part subdivision within a hierarchical classification of things.

Farradane does not develop this classification in enough detail for it to be applied. Nevertheless, a handful of key lessons can be drawn:

> Principle 3: It makes sense to identify key subclasses of causal relationships. Among these will be physical activities, physiological/biological activities, and mental activities. In all three cases, both relationships and self-activity can be identified.
>
> Principle 4: Some relators can be conceived as combinations of more basic relators. This insight can potentially simplify the classification of relationships, and particularly the notation required for this. As was noted above, linked notation can allow us to readily distinguish acting from reacting, and necessary from sufficient from neither, while allowing us to capture many causal verbs as combinations of verbs.
>
> Principle 5: Casual empiricism suggests that some activities (moving) will appear much more often than others (breathing) in a general classification. It is likely desirable to provide simple notation for a small number of often-used relationships, and more complex notation for a larger set of more rare relationships. Such a practice should also facilitate the use of combined notation for some complex relationships, if it is the more commonly used relationships that are most often combined.

These insights suggest that a classification of relationships may resemble a classification of things in important ways. Yet the idea of hierarchy is less intuitive when applied to verb-like relationships than when applied to noun-like things. Fellbaum (2002) suggests that we should think of verb subclasses as primarily expressing 'manner': running and flying involve moving in a particular manner. In some instances, manner can be expressed along one or more continua: 'walk versus run' capture speed of movement, while 'slam versus chop' capture intensity of hitting. Fellbaum also appreciates that verb subclasses might express differences in function (criticized versus advised as subsets of talking) or result (win versus lose as subsets of playing). She does not seem to have appreciated the possibility of whole/part subclasses (basting and rolling are subsets of cooking). Fellbaum also suggests that verb hierarchies are flatter than noun hierarchies, extending three or four levels at most. Importantly, she found that subordinate verbs often do not behave in the same way as their superordinate verb. In other words, they are not true subordinates. This suggests that we should be open to the possibility of using compounding (of verbs with other verbs, adverbs, or things) rather than hierarchy if this provides increased clarity. Fellbaum's suggestion that subordinate verbs often express manner further justifies this approach. Fellbaum found that different types of manner mattered for different types of verb: speed and type of transport for move verbs, force for hit verbs, type of fight for fight verbs, purpose and means for communication. She also found that direction of movement is generally far more important than manner. This too can best be captured by compounding a verb with some indicator of direction or location.

Fellbaum implicitly provides several other arguments for the use of compounds. She notes that exercise has a similar relationship to move verbs as pet has to animal nouns: it expresses function (running for exercise). Punish has few subordinate terms, but is associated with types of hitting, prison, etc. Wave, nod, and shrug are movements of particular body parts but are more importantly gestures. Finally there is a set of result verbs including open, shut, melt, break, destroy, and clean. Fellbaum says they can have subordinates too (slam, bang). But again these seem more like compounds than subordinates.

We can thus state two more principles:

> Principle 6: Compounds of verbs will often prove superior to a hierarchical classification of verbs. In particular, 'manner' may best be captured through compounding. We should strive toward schedules of 'basic verbs' that can then be combined to generate a much larger set of more complex verbs.
>
> Principle 7: Compounds with things (such as locations or directions) will also prove important.

A creative writing instructor once told me, "If you are using an adverb, it is because you are using the wrong verb." While creative writers may benefit from the fact that most verb/adverb combinations can be represented by another verb, the classificationist will want to place these synonymous meanings

within a single subclass. This insight also accords with Fellbaum's advice regarding manner:

> Principle 8: Use compound terms of basic verbs and adverb-like qualifiers whenever possible to capture the meaning of a complex verb.

If a classification will rely on post-coordinated synthetic terms, then it is essential that there always be one obvious way to make a compound. The classifier should not face a choice between 'ruffle feathers' and 'cause feathers to be ruffled.' This obvious point has an important implication: that we would want the concept 'ruffle' to appear only once in the classification. Since classifications of things and properties have progressed farther than classifications of relators, and our purpose in this paper is to achieve a concise yet comprehensive classification of relators, a pragmatic principle follows:

> Principle 9: When a concept has applications as relator and/or thing and/or property, it should generally be classified as thing or property, and then the relator will generally be captured by the compound 'causes x.'

5.0 Inductive Insights

We have still only scratched the surface of a detailed classification of relationships. We have identified perhaps some twenty relationships so far. What other relationships might be useful? Given that deductive efforts have taken us only so far, it makes sense to ask what induction might achieve at this point. Several different source of inductive insight are explored here.

5.1 Surveying Dewey

One way forward is to look at the sorts of causal relationships that are implied by the subclasses that exist within existing enumerative classifications. While there may have been various biases at work in the construction of existing classifications, such an approach should nevertheless at least potentially identify a much larger set of relationships. I have pursued such a research project with respect to the early social science classes in the Dewey Decimal System (DDC) (see Szostak 2011b, 2011c). Over this set of subclasses the following set of relationships was found to be particularly useful:

Influence (Causation/influence in one or both directions). Many subclasses are defined only in terms of 'influence' rather than any particular type of influence.

Controlling/ supervising
Deciding (about)
Evaluating, judging
Moving
Paying/financing
Selecting from
Talking to

It was also found useful to designate:

does not cause/ influence
conflict

Many further relationships can be generated by combining these. For example:

'not' and 'control' means disobedience; combined with 'moving' means 'disobedience of type violent'
'control' and 'paying' means 'control by incentives'
'control' and 'talking' means 'persuasion'

Finally, but importantly, it also proved necessary to classify certain 'changes' that may occur within a particular phenomenon.

growth/development
decline
fluctuations/cycles
stability of
collection of, number of

It should be stressed that this very small list of relationships proved to be sufficient to translate the vast bulk of this substantial number of DDC subclasses (roughly from 300 to 340 in DDC) into compound notation combining these relationships with one or more things in a format such as (thing)(relationship)(thing). This result, though tentative, indicates strongly that a manageably small classification of relationships can be devised that would be of great use to classificationists.

5.2 Khoo's Classification of Verb Types

Khoo (1995) examined all verb entries in the *Longman Dictionary of Contemporary English* 2nd ed. He identified causal verbs according to the criterion that they be transitive verbs that specify the result of some

action, event or state, or express the influence of some object (Khoo 59). While this list may exclude some verbs that we would wish to include—verbs such as mar, surround, and marry are excluded because the subject of the verb is inseparable from the verb itself— it nevertheless provides a very extensive inductive source of causal relationships. Khoo's list is particularly valuable because he employed simple decision rules both in identifying causal verbs and in grouping these; he was not motivated by a desire to establish a verb hierarchy. He does nevertheless identify four broad groups: verbs that indicate cause, verbs that indicate being caused by, verbs that indicate prevention, and verbs that describe some effect.

Among the first (and by far the largest) group, he identifies 64 subclasses. To preserve space, these are listed in the Appendix. For the most part, Khoo's subclasses are logically distinct. In one case, a subclass that subsumes verbs for 'become more active' or 'return to life,' calls for subdivision. In a handful of cases, there is an overlap between subclasses such as between 'cause (something) to have a different shape' and 'cause to have a physical feature.' Note that both of these subclasses can be addressed best through compounding the verb 'cause' with entries from a list of shapes or features. This is true for many of Khoo's subclasses.

Khoo then provides two sets of miscellaneous verbs that do not fit neatly into any of the above subclasses. The first set involves verbs 'where the effect can be described with an adjective.' All of these verbs are compounds of 'to become' (in a minority of cases simply 'to be') and can be easily dealt with by combining 'to become' (or 'to be') with a noun or adjective.

The second set of miscellaneous verbs does not lend itself to such a simple solution. Yet compounding with nouns or adverbs is again often useful. And despite their entry in a miscellany class many of these verbs have synonyms that can be placed in one of the subclasses mentioned above: 'acclimatize' means something like 'to become familiar or comfortable,' 'afforest' means 'to cause to become forested,' and 'advantage' means 'to achieve some priority.' There are only thirty-two cases where neither a synonym nor compound suggest themselves fairly immediately (see Appendix).

Khoo's other three classes are much more readily addressed. Khoo's second main class of verbs, which contains only three entries, includes those that mean 'to be caused by.' If compound terms will be used in a classification, then 'A causes B' should be indistinguishable from 'B is caused by A.' Thus, no special attention need be paid to this class.

Khoo's third main class is verbs that mean 'to prevent.' The subclasses are:

Verbs that mean 'to prevent an event, or the existence of a thing.'
Verbs that mean to prevent someone from doing something
Verbs that mean to persuade someone not to do something

Finally, Khoo's fourth main class involves verbs that indicate some effect, but do not specify its nature (such as 'affect'). As discussed above, it may not be necessary to distinguish such verbs from 'to cause' verbs, though perhaps some indication of 'strength of causation' is called for.

Our analysis of Khoo suggests that we may need some 100 distinct verb relators at most. Compounding will allow this number to be substantially reduced.

5.3 Natural Semantic Metalanguage

Another useful potential source for induction is the literature on Natural Semantic Metalanguage (NSM). A network of linguists is striving to identify common building blocks across all languages. To my knowledge, the implications of this research for classification have not previously been explored: "The method has applications in intercultural communication, lexicography (dictionary making), language teaching, the study of child language acquisition, legal semantics, and other areas" (Goddard, n.d.). "The basic idea of the NSM approach is that we should try to describe complex meanings in terms of simpler ones. For example, to state the meaning of a semantically complex word we should try to give a paraphrase composed of words which are simpler and easier to understand than the original" (Goddard n.d.). Some of these more basic words will be things or descriptors, and others will be relationships. These 'semantic primes' (e.g., do, because, good, you, something, know) have conceptually simple meanings expressable in all languages but are essential to understanding more complex words and phrases.

This approach is obviously congenial to the principles of compounding expressed above (and in Szostak 2011c). That being said, a useful classification of concepts or relationships need not burrow down to the level of absolute primitives sought by

NSM but can be satisfied with words that carry very similar meanings across contexts. It is noteworthy in this regard that NSM scholars speak of 'molecules' or compounds that carry similar meanings across languages (See also Green 2002).

Goddard (n.d.) produces Table 1 (see Appendix), a categorization of the semantic primes identified to date. This table is much less extensive than the list generated by Khoo, and is largely a subset of Khoo's terms. It thus does not much complicate our enterprise, while providing strong justification for our emphasis on compounding.

5.4 The SUMO Ontology

The development of ontologies in the twenty-first century bears some resemblance to the development of classification systems over a century ago: many competing systems exist grounded in conflicting principles. The ontology that seems to have engaged relationships in the greatest detail is the Suggested Upper Merged Ontology (SUMO). Notably for our purposes, SUMO has attempted an exhaustive mapping of every relationship identified in the ontology to every set of verbs (verb synset) identified in WordNet. The SUMO ontology thus provides a very comprehensive list of potential relators. Notably, the ontology is not designed specifically for knowledge organization, but is rather an attempt at making common language machine-readable. It provides logical definitions of each of its terms, and aspires to be both precise and language independent (SUMO has been successfully mapped to many languages; its creators are confident there is no structural barrier to doing so). The ontology contains nouns, adjectives, verbs, and adverbs. Note that the SUMO ontology does not itself identify compound terms, though the purpose of the ontology is to allow complex human utterances to be translated into terms defined within the ontology.

SUMO suggests subclasses for many of its main classes, but these are far from fully developed. It is necessary to also look at the WordNet synsets that it maps major sub-classes onto, and ask how the (often diverse) set of terms found there might best be handled within a classification of relationships. Since SUMO claims to map onto all WordNet synsets, this approach potentially ensures that all verb synsets are placed somewhere within the classification. (The author has thus not directly surveyed WordNet nor other linguistic databases such as VerbNet [which has identified 274 main verb classes]; note that none of these databases take compounding as a primary focus of analysis.) Of course, space does not allow us to list each of the thousands of verbs, even in the Appendix. Rather, close synonyms will be ignored, though at times questions will be raised as to whether certain near-synonyms require special treatment. Most often, compounding (for example with an adverb denoting intensity) provides the best means of coping with near-synonyms.

By following the analysis into the WordNet synsets we inevitably end up with some hundreds of distinct verb subclasses, a much larger set than those identified by Khoo. Again, though, opportunities for compounding abound. The SUMO ontology also provides further justification for division of relators into physical, biological, and intentional categories, and provides some pointers on further subdivisions.

5.5 The OBO Relation Ontology

The OBO Relation Ontology (2005) provides a list of primarily hierarchical relationships that is now incorporated into most ontologies within the biomedical field. It provides logical definitions of the "is a" and "part of" hierarchical relationships (which extant library classifications regularly violate). It also specifies two subsets of "part of": "integral part of" (meaning critically important part) and "proper part of" (which implies distinctness). There are also three locational relations: "located in," "contained in," and "adjacent to." There is also "transformation of," "derived from," and "preceded by," "has participant," "has agent," and "instance of" (the last should be captured within a hierarchical classification of things; "has agent" would be captured by "caused by agent"). It is noteworthy that this very small set of relators has been found to be useful across many ontologies.

5.6 The Unified Medical Language System

The Unified Medical Language System is another useful source, for it strives to overcome the terminological differences across the wide variety of biomedical databases used in the world. It includes a semantic network: a set of entities and the relationships between these. McCray and Bodenreider (2002, 185) note that most of the relationships employed have applicability beyond the field of biomedicine. There are two categories of relationships in the UMLS semantic network. The first is the type/kind relationship. The second are associative relationships of five types: physical, spatial, functional, temporal, and conceptual.

Some of the physical relationships capture whole/part relationships. Others denote the physical composition of something; our classification of non-causal relationships needs to allow for that. Finally, some refer to connections (tributaries, branches, interconnected); these also should be captured in non-causal relationships. Spatial and temporal relationships have been discussed above (though place must be found for terms stressed by UMLS such as 'surround' and 'traverse'). Some of the subclass of conceptual relationships captures non-causal relationships: property of, derivative of, developmental form of (developmental forms are properly 'things' but development itself is a relationship), conceptual part of, and issue in (these last two are properly things). The rest of the conceptual relationships, and all of the functional relations, address causal links (see Appendix).

In addition to these explicit relationships, the UMLS also contains implicit relationships within its classification of entities. There are two types of entities: things and events. This latter subclass is comprised largely of terms that are at least in part causal relationships, such as 'Social and individual behavior,' or the 'functions' and 'dysfunctions' of various organs and tissues. Even within the classification of things there is an important set of implicit relationships. Chemical substances are classified both structurally in terms of their chemical composition and functionally in terms of the effects they can have. The latter would best be captured with compounding: 'Certain chemical [that] causes certain effect.'

5.7 The Art and Architecture Thesaurus

Various thesauri could be consulted. For this paper, the *Art and Architecture Thesaurus* (*AAT*) produced by the Getty Museum has several advantages: it has a broad coverage, though with a focus on the arts that is lacking in the sources consulted above; it is itself an inductive exercise driven by suggestions from users; and the hierarchy underpinning the thesaurus is not only explicit but readily accessible (*AAT* n.d.). As with other sources above, the *AAT* has a hierarchy of relationships (in this case an 'activity' facet), but also addresses relationships elsewhere. Indeed, most of the relationships identified by *AAT* occur outside of the 'activity' facet under headings such as 'economic function,' 'communication function,'or 'entertainment.' The Appendix provides the hundreds of distinct entries that can be found under relevant headings in the thesaurus. As with Khoo and SUMO, compounding again provides a way of coping with this diversity. Most obviously a wide variety of artistic tasks can be captured by compounding basic verbs with substances or shapes as in 'apply paint.'

6.0 Concluding remarks

This paper sets the stage for the development of an exhaustive but manageable classification of relationships. Our inductive efforts have yielded a large set of relators, but this number can potentially be reduced dramatically through the use of compounding. Not surprisingly, there is a huge overlap in the terms identified within the different inductive sources. The task of constructing a classification according to the principles identified earlier in this paper will be taken up in a subsequent paper in this journal.

References

Art and architecture thesaurus online. [n.d.] J. Paul Getty Trust. Available http://www.getty.edu/research/tools/vocabularies/aat/.

Barriere, Carolyn. 2002. Hierarchical refinement and representation of the causal relation. *Terminology* 8: 91-111.

Melgar Estrada, Liliana M. 2011. Topic maps from a knowledge organization perspective. *Knowledge organization* 38: 43-61.

Farradane, Jason. 1967. Concept organization for information retrieval. *Information storage and retrieval* 3: 297-314.

Fellbaum, Christiane. 2002. On the semantics of troponymy. In Green, Rebecca, Bean, Carol A., and Myaeng, Sung Hyon eds., *The semantics of relationships.* Dordrecht: Kluwer, pp. 23-34.

Gardin, Jean-Claude. 1965. SYNTOL. New Brunswick, NJ: Rutgers Graduate School of Library Service.

Goddard, Cliff. n.d. *Natural semantic metalanguage*. Available http://www.une.edu.au/bcss/linguistics/nsm/semantics-in-brief.php.

Green, Rebecca. 2002. Internally-structured conceptual models in cognitive semantics. In Green, Rebecca, Bean, Carol A., and Myaeng, Sung Hyon eds., *The semantics of relationships.* Dordrecht: Kluwer, pp. 73-90.

Green, Rebecca. 2008. Relationships in knowledge organization. *Knowledge organization* 35: 150-9.

Green, Rebecca, Bean, Carol A., and Myaeng, Sung Hyon. 2002. Introduction. In Green, Rebecca, Bean, Carol A., and Myaeng, Sung Hyon eds., *The semantics of relationships.* Dordrecht: Kluwer, pp. vii-xvi.

ISKO Italia. 2004-2011. *Integrative levels classification*. Available www.iskoi.org/ilc.

Khoo, Christopher. 1995. "Automatic identification of causal relations in text and their use for improving precision in information retrieval." PhD diss., Syracuse University.

Khoo, Christopher, Chan, Syin, and Niu, Yun. 2002. The many facets of the cause-effect relation. In Green, Rebecca, Bean, Carol A., and Myaeng, Sung Hyon eds., *The semantics of relationships.* Dordrecht: Kluwer, pp. 51-70.

Mazzocchi, Fulvio et al. 2007. Relational semantics in thesauri : some remarks at theoretical and practical levels. *Knowledge organization* 34: 197-214.

McCray, Alexa T., and Bodenreider, Olivier. 2002. A conceptual framework for the biomedical domain. In Green, Rebecca, Bean, Carol A., and Myaeng, Sung Hyon eds., *The semantics of relationships.* Dordrecht: Kluwer, pp. 181-98.

Nutter, J. Terry. 1989. A lexical relation hierarchy. Technical report TR 89-6, Computer Science, Virginia Polytechnic Institute and State University. Available http://eprints.cs.vt.edu/archive/00000143/01/TR-89-06.pdf.

OBO Relation Ontology. 2005. Available http://www.obofoundry.org/ro/.

Perreault, Jean M. 1994. Categories and relators: a new schema. *Knowledge organization* 21: 189-98.

Stock, Wolfgang G. 2010. Concepts and semantic relations in information science. *Journal of the American Society for Information Science & Technology* 61: 1951-69.

Suggested Upper Merged Ontology. n.d. Available http://www.ontologyportal.org/.

Szostak, Rick. 2003. Classifying scholarly theories and methods. *Knowledge organization* 30: 20-35.

Szostak, Rick. 2004. *Classifying science: phenomena, data, theory, method, practice*. Dordrecht: Springer.

Szostak, Rick. 2011a. *The Basic Concepts Classification, Version 2*. Available http://www.economics.ualberta.ca/en/FacultyandStaff/~/media/economics/FacultyAndStaff/Szostak/Szostak-Basic-Concept-Classification2.pdf.

Szostak, Rick. 2011b. *Translation table: DDC to Basic Concept Classification. Available* http://www.economics.ualberta.ca/en/FacultyandStaff/~/media/economics/FacultyAndStaff/Szostak/Szostak-Dewey-Conversion-Table.pdf.

Szostak, Rick. 2011c. Complex concepts into basic concepts. *Journal of the American Society for Information Science and Technology* 62: 2247-65.

Szostak, Rick, and Gnoli, Claudio. 2009. Beyond aboutness: classifying causal links in the service of interdisciplinarity. In *Proceedings of ASIST Special Interest Group on Classification Research 20th Workshop, Vancouver, November 7, 2009***.** Available http://hdl.handle.net/10150/105534.

VerbNet. [n.d.] *Martha Palmer | Projects |* Available http://verbs.colorado.edu/~mpalmer/projects/verbnet.html.

The Trustees of Princeton University. 2011. *WordNet: A Lexical Database for English*. Available http://wordnet.princeton.edu/

Zeng, Marcia Lei, flumer, Maja, and Salaba, Athena. 2010. Functional requirements for subject authority data (FRSAD). International Federation of Library Associations and Institutions. Available http://www.ifla.org/files/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf.