Learning Idioms – With and Without Explanation

Uri Zernik
Computer Science Department
University of California
Los Angeles, California 90024

Current Address:
GE Corporate Research and Development
Schenectady, New York 12301

Abstract

Modeling learning in any domain can be pursued in two alternative directions: either by finding domain-specific heuristics, or by applying general machine-learning methods. In the linguistic domain, programs such as FOUL-UP [1] and CHILD [2] have used specific heuristics, regardless of the general learning methodology; other programs [3,4] have focussed on the general learning methodology. In our particular task–learning idioms from examples, the evaluation of general learning models is very appealing, since such methods have been studied extensively, and may offer ready solutions.

Language is a special domain in regard to learning. Consider for example behavior of idioms. By definition, idiosyncratic properties are not systematic and are not predictable. Thus, how do people learn such properties from examples? What is the general learning model, therefore, which accounts for idiom acquisition? We examine here the processes involved in acquisition of idioms, and relate them to existing machine-learning paradigms.

1. Introduction

Although idioms are pervasive in human communication, especially in spoken language, their irregular behavior has not been investigated systematically. Idioms are interesting since they provide singular points by which linguistic theories and machine learning theories should be evaluated. Here is a puzzling phenomenon [5,6], which has not been explained so far:

(1) In 1977, Israel and Egypt resolved their long conflict.
The hatchet had been buried.

(2) Finally the patient succumbed.
The bucket had been kicked.

These two phrases (kick the bucket and bury the hatchet) behave differently with respect to the passive voice. While the first paragraph is generally acceptable to native English speakers, the second one sounds awkward. This linguistic behavior is surprising (and unpredictable in computer-program terms) since this pair of phrases are structurally similar. We investigate this idiosyncratic property in terms of its acquisition. This behavior is significant because it sheds light on the otherwise hidden language-acquisition processes and illuminate the role of metaphors.

Metaphors provide learners with the information necessary for explaining the nature of idioms. When encountering an idiomatic phrase, people seek the clues which would make it less arbitrary. However, a metaphor might be obscure, at least from the view of a learner, and thus an explanation cannot always be constructed. Regarding this problem, the questions we consider are: (a) can people learn an idiom even when its metaphor is not accessible? (what is the metaphor underlying kick the bucket?) (b) what is the impact of the metaphor on the use of idioms?

2. The Linguistic Issue: The Passive-Voice Anomaly

Consider the two phrases introduced in sentences (1) and (2). Why can the phrase kick the bucket not take the passive voice? Are there other such phrases? In fact, there are: put one's foot down, for example, cannot appear in the passive (his foot was put down does not convey the meaning of the idiom). How do language learners predict for each new phrase, its behavior, i.e., whether it takes the passive voice or not? There must be a rule for supporting this prediction.

Traditionally, linguistic systems accounted for this phenomenon by including arbitrary syntactic restrictions in the lexicon. Accordingly, lexical entries included explicit clauses to inhibit the passive voice.

pattern: kick the bucket	pattern: bury the hatchet
passive: **not possible**	passive: **possible**

We argue against the inclusion of such arbitrary structural restrictions and seek a more substantial solution in which underlying metaphors account for idiom behavior.

Compare the metaphors underlying our pair of idioms. On the one hand, in bury the hatchet, learners assume that *hatchet* stands for a generic war implement, whose disposal ends a war (similarly, the poet buried her pen could mean that she stopped writing poetry). Some learners can imagine a remote culture in which burying a hatchet is a ceremonial act in signing a peace treaty. On the other hand, no such metaphor could be found for kick the bucket. Learners cannot come up with an explanation as to why the words kick and bucket really mean "to die".

Discourse analysis concerns *why* people phrase utterances the way they do. Sandy Thompson [7] gives a detailed account of people's use of the passive voice. She concludes

that the passive voice is not just an arbitrary feature, but it is a device for accomplishing certain communication goals, such as placing the discourse topic at the beginning of the sentence. In light of this fact, consider the semantic structure of our idioms. Although the reference the hatchet does not stand for any concrete knife, it symbolically refers to the enablement of a fight between two parties. In contrast, the bucket is merely a literal. It does not refer to any object in the context. In the lexicon this difference is denoted as follows:

(3) X:person bury:verb **<Y:the hatchet>**
(4) X:person kick:verb **<the bucket>**

The marked object in (4) is merely a literal; similarly, the marked object in (3) also specifies the surface appearance of the reference. In addition, in (3) there is a variable which can be bound to a concept. Accordingly, a discourse structure can be conceived, in which **Y**, the concept associated with the hatchet, is the topic (as I tried to show in example (1) above), and consequently a communication goal exists for a speaker to apply the passive voice. However, no context can possibly be conceived in which the bucket refers to a topic–since it is merely a literal.

In conclusion, it turns out that the question itself, *can a phrase take the passive voice?* is wrong. The correct question should be: *Is there any discourse structure, in which the phrase should appear in the passive?*. The phenomenon is interesting in terms of machine learning, since it enables us to examine learning concepts with and without explanation, and to show how the availability of explanation (or a metaphor) makes a difference in applying the acquired concepts.

3. Three Machine-Learning Paradigms

What is the method by which a model can *predict* linguistic properties of idioms (do/do-not take the passive voice) as well as their meanings, by being given a small number of examples? Three machine learning paradigms are considered.

In **learning by rote**, an idiom is copied as a chunk. However, this treatment is unacceptable since: (a) idioms must be *generalized* semantically to be applicable in a variety of situations, different from the original one, (b) idioms, as shown by examples (1) and (2), possess their own internal grammar [8], which must also be acquired. Thus idioms cannot be acquired merely as "extended words".

In **similarity-based learning** (SBL), by being given a sufficiently large ensemble of examples, the model acquires surface features which are shared by a number of instances. There is no justification as to *why* a feature is acquired. This approach raises three problems: (a) humans are able to acquire idioms from few, or even from a single example, (b) humans do not require negative examples (e.g., "the bucket was kicked is incorrect"), and (c) humans do not acquire spurious features. Imagine two coincidental instances such as In 1977, X buried the hatchet, and in 1977, Y buried the hatchet. Humans are not thrown off by this co-ocurrence, while an SBL model would assume that In 1977 is a mandatory part of the phrase. Similarly, the appearance of a phrase in the past tense could be taken as a mandatory property by an SBL model.

Explanation-based learning (EBL) remedies these problems by acquiring only features whose appearance can be justified. Spurious similarities are thus ruled out, and significant features can be acquired from a single example. However, this method requires existence of perfect explanatory information, whereas humans are required to learn certain phenomena even without explanatory information. For example, from the learner's point of view, there is no explanation (provided by a metaphor) for the idiom kick the bucket. What is then the reason that the sequence of words kick, the, and bucket means "to die"? Unfortunately–for language learners– many such linguistic phenomena appear arbitrary.

In our model, explanation is a by-product of hypothesis formation; and, accordingly, the existence of a metaphor determines the quality of the explanation. Thus, three questions must be answered:

- How is an explanation constructed for an idiom?
- To what extent can idioms be explained?
- How does explanation effect idiom application?

Following Lebowitz [9] we investigate learning as a function of the state of the explanatory knowledge.

4. The Algorithm

A learning algorithm has been implemented within a program called RINA. In the RINA's phrasal lexicon, an entry, or a *phrase*, is an *association* of a syntactic *pattern* with a conceptual *meaning*. Accordingly, learning is defined as formation of pattern-concept *associations* (unlike other systems [10, 11, 9] which acquire concepts in general). The learning algorithm is *provided* with two knowledge sources: (a) linguistic knowledge and (b) world knowledge.

The lexicon: A partial lexicon is assumed to be given. This given lexicon includes the single words and phrases involved in the input text.

World knowledge: The representation of the explanatory context is assumed accessible to the program. This includes (a) knowledge about the specific domain, and knowledge about other domains alluded to as metaphors.

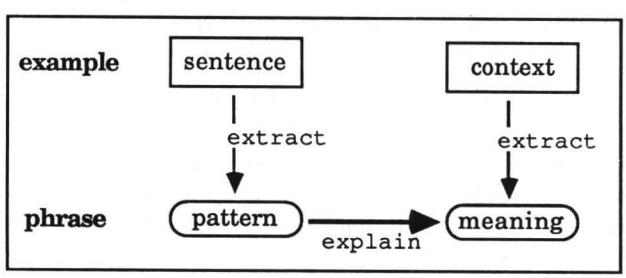

Figure 1: The Learning Algorithm

The algorithm takes as input (a) the *sentence* provided by the user and (b) the explanatory *context*. The inductive leap in learning involves (a) the extraction of the *pattern* from

the context and (b) the extraction of the *concept* from the context. *Explanation* provides the association between the two acquired elements and supports both steps (1) and (2). The construction of the explanation is not a process on its own* as in [10], rather it is a by-product of learning. Moreover, explanation cannot always be constructed. In fact, the point of this paper is to compare cases where explanation can and cannot be constructed.

5. The Process Model

Next, we show the steps in learning bury the hatchet from an explanatory context.

Detecting a Discrepancy: The acquisition of a new phrase is initiated only if there is no other valid interpretation for the sentence. There is an interpretation for the input sentence In 1977, Israel and Egypt buried the hatchet, based on the single words.

> **pattern:** X:person bury Y:phys-obj
> **concept:** X ptrans Y, causing Y to be under the ground

While syntactically intact, this interpretation, which amounts to "in 1977 they put a hatchet under the ground", fails on semantic grounds. Burying a hatchet does not relate to the given goal-plan situation. In fact, the program detects a near miss [12]. The buried object is not some arbitrary physical object, but it is a fighting tool–albeit for a different kind of fight. Thus, a new hypothesis must be formed.

Extracting a Phrase Pattern: Four discrepancy-driven strategies are employed in this case, where the general intuition is to maintain *frozen* elements which cause parsing failures and to *generalize* elements which can be explained by the parser.

(1) The modifier In 1977 is excluded altogether as it is matched by an existing lexical phrase, and it is taken as a standard modifier. (The prepositional phrase In T:year exists as a lexical phrase which is taken as a time modifier.)

(2) Both Israel and Egypt can be resolved in the context; and, therefore, these names are variabilized.

(3) The conjunction itself is not taken as a mandatory element. Conjunction is explained by the context, since Israel and Egypt are both linked by the goal-conflict situation. Thus, the reference becomes:

<Israel and Egypt> —> <N:nation>

(4) In contrast, the reference the hatchet cannot be resolved, and henceforth it is maintained as a literal:

<the hatchet> —> <the hatchet>

Clearly, this process depends on the actual contents of the lexicon (In T:year exists in the lexicon) and actual world knowledge. In the state of knowledge assumed here, the new formed pattern is:

N:nation bury:verb the hatchet

* To avoid confusion, notice that "explanation" is used here to denote a static object and not a process.

Extracting a Phrase Meaning: The meaning is constructed by parsing the user's clarification:

> Israel and Egypt had been involved in a bloody conflict.
> In 1977, they signed a peace treaty.

The first sentence establishes the goal-conflict situation (GC) between Israel and Egypt (I&E). GC can be resolved by two alternative plans: (a) fight (O1), or (b) negotiate (O2). The second sentence is taken as a switch from O1 to O2. Thus, the meaning is taken as:

- goal conflict (GC) exists between Israel and Egypt (I&E)
- I&E quit the fight

At this point, the new phrase has a pattern and a meaning:

pattern: <N:nation> bury <H: the hatchet>
meaning: N is involved in a conflict C
N quits fight F

In contrast to the variable N which is bound to a concept in the meaning, the variable H is unbound. No hatchet was found in the context, so the binding for H will be obtained from the metaphor.

The Explanation: The explanation maintained by a phrase *connects* the phrase *concept* to the phrase *pattern*. This is only a connection and not a proof since, by their nature, metaphors do not fully account for meanings of idioms. In our example, four clauses are used in the explanation obtained by the metaphor:

(1) I&E bury hatchet H % the sentence itself
(2) **if** X bury Y
and Y is a tool
===> X disenable-use of Y % generalizing "bury"
(3) hatchet H ===> war-implement H % generalizing "hatchet"
(4) **if** X disenable-use of implement I for plan P
and X executing plan P
===> X intend to quit plan P % an inference rule

These four given clauses interact in the following sequence:

(5) —> I&E disenable-use of hatchet H % by (1) and (2)
(6) —> I&E disenable-use of war-implement H % by (5) and (3)
(7) —> I&E intend to quit fight P % by (6) and (4)

Proposition (7) matches the meaning of the phrase, so the entire explanation is maintained as a clause in the phrase.

pattern: <N:nation> bury <H: the hatchet>
meaning: N is involved in a conflict C. N quits fight F
explanation:
—> I&E disenable-use of hatchet H
—> I&E disenable-use of war-implement H
—> I&E intend to quit fight F

Consequently, the variable H in the pattern gets its binding in the *explanation* clause (and not in the concept itself).

6. Learning Without Explanation

In learning kick the bucket, no such explanation exists. The pattern and the meaning are both extracted, but the metaphor which could provide an explanation remains obscure.

The constructed phrase is:

pattern:	<P1:agent> kick <P2: the bucket>
concept:	P1 die

The reference P2 is not associated with any concept, since a bucket cannot be found either in the context (there is mention of a bucket in previous discourse) or in the explanation (since the explanation does not exist).

In a different learning situation an explanation **could** be constructed for the same phrase. Suppose for example the following input context:

John was standing on the bucket with a rope around his neck.
Then he kicked the bucket away and fell to his death.

In this case an explanation could be constructed as:

—> X kick the bucket B
—> X is not supported by B
—> X is hanging from the rope R
—> R strangles X
—> X dies

Interestingly, second language learners, when presented with this context, *are able* to generalize the phrase and even use it in the passive voice, since they have an explanation.

7. Conclusions

We have analyzed phrase acquisition as a function of information provided to the learner. We have identified three information sources and evaluated their contribution:

(1) Lexical entries given for words and phrases.

(2) The context underlying the utterance phrases.

(3) Metaphors associated with single words.

Both in parsing and in learning, meanings of word combinations are derived by using the single words. However, in parsing, the derivation of meaning involves simply looking single words up in the lexicon. In learning, on the other hand, the literal derivation is not acceptable, and the process is complicated. In learning, meaning derivation involves a search for metaphors implied by the single words and selection of concepts from the context.

We have explained two surprising phenomena:

(1) How can a phrase be learned from a single example (bury the hatchet)? To that end, we have shown strategies for concept extraction which rely on linguistic and world knowledge.

(2) How can a phrase be learned even when a metaphor does not exist (kick the bucket)? In fact, the same learning strategies as in (1) are employed also in absence of a metaphor.

However, when a metaphor does not exist, the association between the pattern and the meaning remains unexplained. The phrase cannot be used in a general way, and its application is limited.

Acknowledgement

I thank Ron Sumida for helpfull comments.

References

[1] Granger, R. H., "FOUL-UP: A Program That Figures Out Meanings of Words from Context," pp. 172-178 in *Proceedings Fifth IJCAI* , Cambridge, Massachusets (August 1977).

[2] Selfridge, Mallory G. R., "A Process Model of Language Acquisition," 172, Yale University Department of Computer Science, New Haven, Connecticut (May 1980). Ph.D. Dissertation.

[3] Reeker, L. H., "The Computational Study of Language Learning," in *Advances in Computers*, ed. M. Yovits M. Rubinoff, Academic Press, New York (1976).

[4] Anderson, J. R., "Induction of Augmented Transition Networks," *Cognitive Science* **1** (1977).

[5] Chafe, W. L., "Idiomaticity as an Anomaly in the Chomskyan Paradigm," *Foundations of Language* **4** (1968).

[6] Dong, Quang Phuc, "The Applicability of Transformations to Idioms," in *Proceedings Chicago Linguistic Society* (1971).

[7] Thompson, Sandra, "The Passive in English: A Discourse Perspective," in *In Honor of Ilse Lehist, Ilse Lehist Fuhendusteos*, ed. Robert and Linda Shokey, Foris (1987). to appear.

[8] Fillmore, C., P. Kay, and M. O'Connor, *Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone,* UC Berkeley, Department of Linguistics (1987). Unpublished Manuscript.

[9] Lebowitz, M., "Integrated Learning: Controlling Explanation," *Cognitive Science* **10**(2) (1986).

[10] Mitchell, T., R. Keller, and S. Kedar-Cabelli, "Explanation-Based Generalization: A Unifying View," *Machine Learning* **1**(1) (1986).

[11] DeJong, G. and R. Mooney, "Explanation-Based Learning: An Alternative View," *Machine Learning* **1**(2) (1986).

[12] Winston, P. H., "Learning Structural Descriptions from Examples," in *The Psychology of Computer Vision*, ed. P. H. Winston, McGraw-Hill, New York, NY (1972).

[13] Zernik, U., "Strategies in Language Acquisition: Learning Phrases from Examples in Context," UCLA-AI-87-1, LA, CA (1987). Ph.D. Dissertation.