Context in Translation: definition, access, and teamwork

Appendix 3

Three Distinctions:
General/Domain; Dynamic/Frozen; and Saying/Said

The purpose of this appendix is to support the claim, made in the body of the article, that non-text is a valid aspect of context. Some would argue that all context can be fully expressed in text of some kind. This appendix begins with a discussion of the distinction between general language and domain-specific language. Then, another distinction is introduced: a distinction that is orthogonal to the general-domain distinction, namely, the dynamic-frozen distinction. Finally, an aspect of the work of the philosopher Levinas is discussed: the distinction between the "saying" and the "said". The saying/said distinction is the theoretical basis for the non-text aspect of context in this article. This appendix is adapted from a paper presented at the January 2010 IGCL conference in Hong Kong.

1. General/Domain

Dividing instances of lexical items in a text between general language and domain-specific language tells only part of the story. The general vs. domain distinction is not always clear. Most people who own a computer know what a disk drive is. The term "disk drive" is part of the domain of computer hardware but has also become an expression in general English. Should the lexical items related to everyday activities such as baking, driving, and sports be treated as terms within domains? A number of years ago, Doug Lenat of the Cyc project (http://www.cyc.com) said (private communication) that what is commonly called general language is the union of about twenty domains and that there is no such thing as general language. Nevertheless, if one includes function words and examines what can reasonably be expected to be understood by speakers of a language with no specialized training, there is a useful distinction between general language and domain-specific language along some continuum. In particular, this distinction is useful when a group of specialists in a technical non-culture-specific domain, crossing national and cultural boundaries, agree on definitions of concepts and terms for those concepts in various languages. In this case, people create domains that include language-independent concept systems to facilitate technical communication. This contrasts with very general language, in which there are few if any truly language-independent concepts. At the general-language end of this general-domain continuum, languages are tied to a culture rather than to a domain. At the international-domain end of the continuum, concepts (e.g. those of inorganic chemistry) are culture independent. In the middle of the continuum, there are concepts, such as national laws, that are domain-specific yet tied to a culture.

2. **Dynamic/Frozen**

There is another distinction relevant to language resources: the distinction between dynamic and frozen language.

A particular usage of a lexical item in a language is dynamic or frozen at a point in time relative to a group of speakers of that language. Strictly speaking, a lexical item, by itself, is not frozen or dynamic, since it could be frozen in one sense and dynamic to someone in another sense. That is, the meaning of a lexical item in a particular situation for a particular person could be freshly created in the mind or retrieved from a frozen state stored from previous situations, but in this discussion, *lexical items* rather than concepts will be treated as dynamic or frozen, by virtue of their current relation to a meaning in the minds of most people in a linguistic community.

General language lexical units can be either dynamic (that is, fresh in meaning to most people, even though each word in the unit is familiar in another context) or frozen (that is, associated with a familiar meaning for most people), as can domain-specific lexical items.

Some lexical items, such as “home equity loan”, are so well known that they are frozen for nearly everyone who lives in the United States. They are frozen in the sense that they are frequently used within a domain, in this case, finance, with a consistent meaning that is retrieved from memory of typical usage. A home equity loan is a loan using a home as collateral when one owes less in mortgages than the value of a home. Other items, such as “underwater” and "upside down", in the recently created sense of owing more than the current value of a home, are dynamic in that the meaning is still fresh to most people. Over time, this metaphorical meaning of underwater and upside down may become frozen, and most speakers will take no special notice of them.

Consider the neologism “screenager”. When I recently encountered this lexical item for the first time, I wasn't sure what it meant. One definition is “a person in their teens or twenties who has an aptitude for computers and the Internet”. As of November 2009, screenager got 169,000 Google hits. Or consider “frankenfood”, which is a somewhat recently coined derogatory word for a food that contains genetically modified ingredients. As more and more people get acquainted with the word frankenfood it transitions from dynamic to frozen. With over 89,000 Google hits, frankenfood, may soon enter into general language.

Dynamic lexical items are not limited to single words. Recently, I saw a bumper sticker on an automobile that read “gun control means using both hands”. At first it made no sense to me. As an advocate of placing more controls on the sale and possession of handguns, I was searching for a legislative meaning. Then, I realized that the expression was devised by someone who is against increased

controls on guns and that it may mean that the only control needed on guns is that if you need to use one, you should hold it with both hands for increased accuracy. I then showed the sticker to two other people. One looked puzzled until the other one said, “I've seen that before” and explained what it meant. At the same instant in time, the expression was fresh to one person (that is, its meaning was created on the spot) and familiar to the other (that is, retrieved), placing it somewhere between dynamic and frozen.

An observation at this point about the dynamic-frozen distinction is that most language resources, whether focused on general language, such as WordNet (Miller 2010), or domain-specific language, such as termbases, are about language that is frozen from the perspective of the person compiling the resource. Bi-text corpora are, by definition, about correspondences between segments of source text and segments of target text that have been frozen by a translator through the act of translating.

Language resources are an excellent way to document or discover something about frozen language. They do not, however, tell a human or a computer how to deal with dynamic language.

Many words have multiple possible meanings. This makes them superficially ambiguous (Melby 1995:55) but not dynamic, as long as a set of allowed meanings can be pre-established.

Traditional rule-based machine translation reached a plateau in performance, in my opinion, because it tried to make frozen language into dynamic language and process dynamic language using simple algorithms. That statement perhaps deserves further explanation. Rule-based machine translation systems usually treat multi-word expressions as artificially dynamic to the system, even if they have been translated many times before as lexical units rather than isolated words.

For example, a banana split is a dessert that includes much more than bananas. A machine translation system, without human intelligence and without the term "banana split" as a unit in its lexicon, would treat this term of American culture as dynamic language. That is, it would process the two words "banana" and "split" separately, probably as two nouns, and would not be likely to generate in the target language a lexical item that would help the reader visualize three scoops of ice cream, a banana that has been split down the middle, and various toppings, including whipped cream, nuts, and a maraschino cherry.

The problem is that the rules in the system are incapable of predicting how a given lexical unit is typically translated by a professional human translator. Data-driven machine translation avoids the assumption that most lexical units are dynamic language by effectively searching through a bi-text corpus for instances of a lexical unit and noting how it has been translated in the past. However, data-

driven machine translation cannot deal with truly dynamic language, that is, expressions whose translations are not found in the bi-text corpus and cannot be adequately translated as a bottom-up composition of individual words.

The dynamic-frozen distinction is independent of the general-domain distinction. That these two distinctions can vary independently results in four categories of lexical units: dynamic general, frozen general, dynamic domain-specific, and frozen domain-specific.

It may be that many texts that need to be translated in today's world consist almost entirely of lexical units that are frozen, that is, found repeatedly in a large bi-text corpus with a consistent sense. However, there will always be the need to deal with dynamic language. Language resources are about frozen language. Dynamic language is about personal interaction with language resources.

3. Saying vs. Said

The dynamic-frozen distinction is about the nature of personal experiences between people and lexical units in texts. The “saying” vs. the “said” in the philosophical writings of Emmanuel Levinas (Critchley 2002; Smith 2005) goes beyond the dynamic-frozen distinction and enters into the realm of the interaction between two people rather than between a person and a text.

Levinas has proposed that when two people are engaged in a face-to-face discussion where each person sees the other party as “other”, that is, as a person not reducible to a controllable object, then each of their responses is a “saying”, that is, a living second-person action. On the other hand, once interaction between people is recorded and played back or re-stated, it becomes a “said”, that is, a linguistic object that can be analyzed from a third-person perspective.

The saying is in the ethical realm and is not tied to any text. The said is ontological in the philosophical sense of exploring the nature of being and is the focus of language resources. The ethical realm certainly includes the "call" of the other for basic needs: food, shelter, safety. In the present discussion, it is assumed that the call of the other also includes the need for mutual intelligibility.

The saying-said distinction has many ramifications. Levinas contrasts ethical behavior, in which one responds willingly to the needs of the other, with unethical behavior, in which one treats another in a “totalizing” manner, that is, as an object to be controlled as part of a panoramic field that can be fully comprehended as a formal system. For Levinas, ethics cannot be reduced to a set of formal rules that can be represented as a text, that is, a said.
Thus, the saying, which involves living interactions in which at least one person is striving to be intelligible to the other, cannot be fully reduced to a text.

The saying-said distinction can be applied to human communication to better understand how people can sometimes speak “past” each other when they do not accurately or sufficiently take into account the freshness of the constantly changing state of the other. They treat what someone else says as a said by merely processing words that can be analyzed and fully understood without engaging in a true conversation.

A central philosophical issue of concern here is whether the second-person perspective of the saying can be adequately modeled by the objective, third-person said. If you disagree with the saying-said distinction and believe that all human communication can be reduced to a mechanistic, third-person perspective, then you might be favorable to Kurzweil's plan to overcome death by transferring the exact state of your brain to a powerful computer (Kurzweil 2000:101-132). If, on the other hand, you agree with Levinas and believe that the saying cannot be captured by a third-person perspective, then you might be inclined to expect limits to what a computer can ever accomplish relative to the sayings that are part of human relationships.

I am not claiming to be able to prove that Levinas is right or wrong. Rather I am claiming that it makes a difference whether he is right or wrong. It makes a difference to what computers can theoretically be expected to do, relative to such tasks as understanding and translating texts.

Theoretical computer science has established that all formal languages of a certain power on the Chomsky hierarchy (Chomsky 1959) are equivalent to an imaginary machine named after Alan Turing, the great computer scientist. The Turing Machine was described (Turing 1936) well before the first general-purpose digital computer was built in the late 1940s, but is still relevant.

Current computers are no more powerful than Turing machines (Church-Turing 2009). Even the more advanced architectures in Flynn's taxonomy (1972), though much more efficient than a traditional von Neuman architecture, are faster but not more powerful than a Turing Machine. That is, current computers can only do that which can be completely captured in a step-by-step fashion, deterministically or non-deterministically, from a third-person perspective. They can only deal with the said.

Levinas claims that a saying cannot be reduced to a said without some loss. The said is always a paraphrase of the saying. If he is right that a saying cannot be completely modeled from a mechanistic, third-party perspective, then it may well be the case that current computers, no matter how powerful, will never successfully deal with all aspects of human-to-human interaction, including the ethical aspects. The saying is beyond the reach of a Turing Machine. Thus, a computer that deals with the saying would have to go beyond the power of a Turing machine in a fundamental way. The human mind, which can engage in the saying, does go beyond what a Turing Machine can do. Levinas did not

discuss his philosophy in terms of computers. The claim that the human mind can do what a Turing Machine cannot do is based on an interpretation of Levinas, who, unfortunately, has passed on and is not available to discuss these questions face to face

When humans read texts or hear stories or see movies about human interaction, they are capable of generating sayings in their minds, as if they had participated in the event described. Even great authors cannot capture the saying in a text, such as a novel, but they are able to inspire many readers to generate sayings. Often, a person will generate different sayings at different periods in their life from the same text, depending on an infinite variety of possible factors, including attitude and previous experiences.

The said, such as a written description of an event, is lacking the kind of ethical dimension that Levinas writes about, the very personal call of the other. This ethical dimension can be precisely relevant to what is an appropriate translation of what is said. We will give a few examples.

A written account of a person asking someone else to do something or apologizing for doing something might be translated differently, depending on the reconstructed saying that goes beyond the written account. Tone of voice and intent behind the tone, in the re-constructed saying, can influence the choice of words with varying connotations. Languages differ in the options available to express human relations and intents. The Japanese lexical system for apologies is more intricate than the English system.

Consider a grandfather who says to his granddaughter, "You look pretty today" vs. a man uttering the same words to a woman at a bar. You might assume a very different intent as you reconstruct a saying in your mind, based on various assumptions. In some languages, the saying could influence the translation. Note that the woman in the bar might be the man's daughter rather than a stranger. This could radically influence the saying.

Interpreters during the Iraq war often reported having difficulties with the words given to them by American military personnel because they felt that a literal conversion of words in Arabic would not help an Iraqi listener understand the intent of the American. They had to carefully decide how to help Americans and Iraqis have the best chance of engaging in a shared saying. Sometimes, more is needed than careful selection of words. A gradually developed understanding of culture is needed as well.

The bottom line is that non-text, as described in the body of the article to which this is an appendix, includes the saying. The saying cannot be reduced to the said without loss, yet it can influence a translation. Thus, non-text is an essential category of context.

The body of the article explains the relation between non-text and other aspects of context..

References

.

Chomsky, N. (1959). "On certain formal properties of grammars". *Information and Control* 2 (2).

Critchley, S. (2002). "Introduction". In S. Critchely and R. Benasconi (eds.), *The Cambridge Companion to Levinas*, (pp. 1-31). Cambridge: Cambridge University Press.

Flynn, M. (1972). Some computer organizations and their effectiveness. *IEEE Transaction on Computing*, *C-21*, 948.

Kurzweil, R. (2000). *The age of spiritual machines: When computers exceed human intelligence.* New York: Penguin.

Melby, A. K. (1982). Multi-level translation aids in a distributed system. In J. Horecky (ed.), *North Holland Linguistic Series #47--COLING82* (pp. 215–220), Amsterdam: North-Holland Publishing Company

Miller, G. A. (2010). A Lexical Database for English. Accessible from: http://wordnet.princeton.edu/

Smith, M. B. (2005). *Toward the outside*. Pittsburgh: Duquesne University Press.

Turing, A. M. (1936). "On computable numbers, with an application to the Entscheidungs problem". *Proceedings of the London Mathematical Society*, *2* (42), 230–265.

Church-Turing (2009). Church-Turing Thesis. In the on-line Stanford Encyclopedia of Philosophy. Retrieved from http://plato.stanford.edu/entries/church-turing/