On the Spanning Hypothesis for EDI Semantics*

Steven O. Kimbrough
University of Pennsylvania
kimbrough@wharton.upenn.edu

Scott A. Moore
University of Michigan
samoore@umich.edu

Abstract

What EDI needs is a good semantics, that is, a workable formal theory of what EDI messages mean. As is widely recognized, the point applies to electronic commerce and to communications by artificial agents in general. Some progress has been made in this direction, but very much work remains to be done. In this paper we introduce and discuss the spanning hypothesis *for agent (or, e.g., EDI or electronic commerce) communication languages. The spanning hypothesis is a claim about the semantics for a given communication language, and we think it represents a necessary condition for truly successful artificial communication in electronic commerce. After discussing and clarifying the hypothesis, and how it might be confirmed, we present evidence in its favor from an analysis of several EDIFACT transaction sets.*

1. Introduction

EDI (electronic data interchange), electronic commerce generally, societies of artificial agents—all these require and use formal communication languages for messaging among computer programs. This fact immediately raises the question of how best to design the languages to be used for messaging. Many such languages exist, are in daily use, and are undergoing refinement (notably X12 and UN/EDIFACT in EDI and electronic commerce, KQML for artificial agents; we refer to all of these as ACLs—agent communication languages). Strong criticisms have been lodged against the designs of these ACLs, especially the EDI languages (see [1, 2, 3, 4, 5, 6]). Most of these criticisms are based on the fact that the theory of what the languages need to say, indeed the theory of what they already do say, is but embryonic. Such a theory is what we call a *semantics* for the language(s) in question.

The need for a semantics for EDI and for ACLs in general is well recognized. The following comment is representative of the literature's consensus.

> Systems based on distributed agent architectures require an agent communications language having a clearly defined semantics. ... ...Without one, agent designers cannot be certain that the interpretation they are giving to a "performative" is in fact the same as the one some other designer intended it to have. Moreover, designers are left unconstrained and unguided in attempts to extend the set of communication actions, and they cannot be certain how the existing or newly defined communication acts will be used in interagent communications protocols. [7]

We agree with this description of the state of affairs today for ACLs. Although further documentation of the scene today would be useful, our aim here is to report progress in solving the design problem for ACLs. To that purpose, let us now consider more carefully why a semantics is needed for an ACL, particularly for one used for EDI.

2. On ACL Design Criteria

We may say that any ACL will consist of a lexicon plus rules of expression formation. (There are important other elements, but they need not concern us here.) The lexicon is like a dictionary for a natural language, the rules of formation like a grammar. Together they (lexicon and rules of formation) determine every possible expression in the (formal, ACL) language in question. They thus set the expressive power of the language. Given this, what do (or should) we require of an ACL when we design it? Perhaps what is most important is that we want the language to be rich enough to say what needs to be said for the application we have in mind. For example, an EDI standard that could not express a purchase order or an invoice could not be taken very seriously. If an ACL can say what needs saying for a given application, we say that it *spans the domain*; it covers what needs to be said.

But spanning is not enough. We also require a degree of parsimony. For example, an ACL with an infinite number of (unspecified) predicates and functions is entirely unworkable from a practical standpoint. Such a language (think of it as all of first-order logic) could span any domain. Whenever expressivity problems were (thought to be) encountered one

*File: span-sok-sam-c.tex. Date: 19980930. Kimbrough's Web page is http://opim.wharton.upenn.edu/~sok/. Moore's Web page is http://www-personal.umich.edu/~samoore/. This work was supported, in part, by DARPA contract #DASW01 97 K 0007.

$$
\begin{aligned}
&\exists e' \exists e'' \exists t' \exists t'' \\
&((po(e) \wedge Speaker(e,s) \wedge Addressee(e,r) \wedge \\
&\quad Theme(e,(e'|e'')) \wedge Cul(e,t)) \wedge \\
&\quad (H(e) \quad \leftrightarrow \quad (delivering(e') \wedge Agent(e',r) \wedge Goal(e',s) \wedge Theme(e',g) \\
&\qquad\qquad Sake(e',e) \wedge Unit(e',g,u_g) \wedge Quantity(e',g,q_g) \wedge \\
&\qquad\qquad Cul(e',t') \wedge t' \leq t_{e'})) \wedge \\
&\quad (H(e) \quad \rightarrow \quad (promising(e) \wedge \\
&\qquad\qquad (K(e) \leftrightarrow (paying(e'') \wedge Agent(e'',s) \wedge Goal(e'',r) \wedge \\
&\qquad\qquad Theme(e'',\$) \wedge Sake(e'',e') \wedge Unit(e'',\$,u_\$) \wedge \\
&\qquad\qquad Quantity(e'',\$,q_\$) \wedge Cul(e'',t'') \wedge t'' \leq t_{e''}))))
\end{aligned}
$$

Figure 1: Representation of a Simple Purchase Order (extensional approximation). (Note: Terms without quantifiers, e.g., e, $, are names, or constants. $(x|y)$ is the compound event consisting of events x and y.)

could simply add new predicates and functions. This trivializes spanning and leads to the sort of unhappy situation we have today with EDI: an undisciplined (and largely uncollected) lexicon with *ad hoc* additions, greatly complicating participation in electronic commerce.

There are other design criteria for ACLs, but these two will do for now. What emerges is a realization that spanning and parsimony are conflicting goals, so that good design requires wise tradeoffs. We might hope for designs that span and yet are maximally parsimonious. In search of this, we could fix the logic (grammar, or rules of formation) for a language and then seek a finite, even small, lexicon that could span the domain. Indeed, how else are we to imagine building an ACL that could really be used by artificial agents? Assuming this strategy, then, the importance of a semantics—a theory of what meanings need to be communicated—becomes apparent: with the semantics to guide us, we know what needs to be said and, more importantly, we can circumscribe the language because we know what does not need to be said.

By way of making these remarks more concrete, consider a particular semantic representation for an EDI message: Figure 1. The expression in the Figure uses Kimbrough's semantic theory for speech acts to state (in first-order logic) the meaning of a simple purchase order (see [8, 9, 10, 11] for development of the underlying theory). By way of explanation, suppose that s sends a purchase order to r at time t. The subject of s's purchase order is g (some goods), which s wishes to purchase in quantity q_g and units u_g, for delivery on or before time $t_{e'}$. For this, s is prepared to pay r an amount of money, $q_\$$, in units $u_\$$ on or before time $t_{e''}$. This is what we need to model with a semantic theory.

Under our previous analysis [12], this purchase order is a sort of speech act, uttered by s to r. The speech act is complex in that it is both a request (for r to deliver the goods) and a conditional promise (that s will pay for the goods, if r delivers them). Given the analysis and the semantic theory, the purchase order is represented as in Figure 1. Is it correct? This question merits—and has had—a great deal of discussion. Notice, however, what we get when we translate Figure 1 more or less literally back into English. Here goes. "e [think of it as the ID on the PO] is a purchase order, from s to r, dated t. A delivery [e'] of goods g is requested for the sake of this purchase order. The delivery should be to s by r (or r's agent), and should occur on or before $t_{e'}$. Measured in units u_g, the amount of goods delivered should be q_g. On condition that the delivery occurs as just described, s promises to pay [e''] r for the delivery. Payment will be in the form of money. Measured in units $u_\$$, payment will be in quantity $q_\$$, and will occur on or before $t_{e''}$."

The match—between the original statement and the English translation of the semantic representation—is, we think, pretty good. Note in particular that inferences we can agree to in ordinary language work logically—formally—when formalized in accordance with this theory. For example, "If s sends a purchase order, e, to r, and r complies by delivering the goods as requested, and s does not then pay r for the goods, s breaks (fails to keep: $\neg K(e)$) s's promise to pay, e, associated with the purchase order." Less stiltedly: "If you send a PO, get the goods, and don't pay, then you have broken a promise."

All this is well and good. It is part of the reasons why Kimbrough's semantics for EDI is a promising enterprise, a contender. Our point in this paper is not to celebrate this fact, but to exploit it. Let us assume that the semantics, albeit very incomplete, is on the right track. What can we make of this?

Rôle	Description
Agent	Volitional initiator of action
Patient	Object or individual undergoing action
Theme	Object or individual moved by action
Goal	Individual toward which action is directed
Source	object or individual from which something is moved by the event, or from which the event originates
Experiencer	Individual experiencing some event
Beneficiary	Object that benefits from the event
Location	Place at which the event is situated
Instrument	Secondary cause of event; the object or individual that causes some event that in turn causes the event to take place
Speaker	Volitional initiator of speech act (cf., Agent)
Addressee	To whom a speech act is directed
Unit	The unit of measurement for a given item
Quantity	The quantity of a given item (in the units)
Cul	culmination time for an event
Sake	x is for the sake of y

Table 1: Examples of thematic rôles. (See also [11].)

3. The Spanning Hypothesis

Consider Figure 1 more carefully. It is remarkable that the expression in Figure 1 uses only predicates from a small controlled vocabulary. Every predicate is either a verb (predicate beginning in lower case) or a *thematic rôle.* Thematic rôles are generic helper predicates used to qualify the meanings of particular verbs. (See [13, 14] for further discussion and their conceptual development in the context of linguistic theory.) Table 1 provides a starter list of standard thematic rôles; we assume the reader is familiar with verbs. What we find especially intriguing about Figure 1 is the possibility that the representational economy generalizes. What if the verbs and the thematic rôles belong to a public, controlled vocabulary; what if they can be well defined; what if they are limited in number? What if the smallish controlled vocabulary can be used in many different kinds of messages; and what if it allows us to express any purchase order, or indeed any EDI message at all? If these questions can be answered in the affirmative, then the strategy we discussed above—of fixing the semantics and identifying a finite, indeed smallish lexicon that spans the domain—will look very promising indeed.

It isn't quite that simple, for two reasons. First, EDI messages *do* need to say things about an indefinitely large number of things: "of shoes, and ships, and sealing wax…" and all the things, tangible and not, that can be subjects of commercial transactions. However, Figure 1—and more generally, Kimbrough's semantics—has something helpful to offer on this point. Notice how things like shoes and ships and sealing wax enter into the representation: through names (or more generally, referring expressions). In Figure 1, g names the (kind of) goods that are the subject of this purchase order. Think of g as the ID number of a catalog item. It is entirely consistent with Figure 1 and the associated semantics for g to be actually given as a document which can be referred to by its URL, or any other expression that uniquely identifies the items that are subject of the purchase order. Thus, we have to qualify the strategy. We can still seek a smallish lexicon of controlled vocabulary, but we must assume that there is an open vocabulary—indefinitely large—of nouns and other referring expressions. This does not vitiate the general programme. Just as finite agents can be told about a smallish lexicon (of possibly complex predicates and functions), so they can be told about a smallish set of catalogs (with possibly very large and changing lists of simple elements).

The second problem is that it is not realistic to expect to fix the controlled lexicon indefinitely. There is always the possibility of some new sort of thing coming along. If that new thing could be defined in terms of existing predicates and functions, then there is no problem. However, sometimes these new things are genuinely new, e.g., a new kind of pollution permit or export license, and new predicates will have to be added to the controlled lexicon. Again, this does not vitiate the programme. First, genuinely new things do not come along all that often. If they did, it is hard to see how automation for electronic commerce could have come as far as it has. Second, Figure 1—and more generally, Kimbrough's semantics—again has something helpful to offer us. Notice how dominant logical conjunction ($\wedge$) is in the Figure. This facilitates graceful, nondisruptive addition of semantic content. Suppose, for example, in the purchase order s wanted to stipulate that the carrier for the delivery should be FedEx. Define a new predicate, *Carrier(x,y)* ("y is the transport company used for event x") and add $Carrier(e', \text{FedEx}) \wedge$ to line 1 of Figure 1. This move generalizes nicely. See [13] for a discussion of this. It yields one of the best arguments for event semantics with thematic rôles (ESΘ semantics).

With this background and these qualifications at hand, we can now state what we think is a particularly interesting and plausible spanning hypothesis for EDI: *Using a semantics and rules of formation of the sort described by Kimbrough (cf., Figure 1) it will be possible to span the domain of EDI (electronic commerce generally) today, using an open vocabulary of nouns (mainly in the form of catalog items) and a smallish controlled vocabulary of predicates, consisting of verbs, thematic rôles, and a number of miscellaneous predicates.*

Of course, this is not a provable hypothesis—but it is testable and it is refutable. On the positive side it can be demonstrated by producing a lexicon that appears to do the trick, and success or failure in that endeavor is what is of primary importance. Of course, our spanning hypothesis is a bit vague—e.g., What do we mean by "smallish"?—but the important thing is to investigate the hypothesis and, e.g., to discover just how large and how constituted the controlled vocabulary should be.

So how can we investigate this spanning hypothesis? One way would be to look carefully at the composition of some existing EDI, say EDIFACT, transaction sets. Is it possible to interpret the elements used to define the EDI messages in a way that fits with ESΘ theory (as in Figure 1)? Can, say, the EDIFACT standards definitions be reduced to a series of predicates consisting of verbs, thematic rôles, and miscellaneous predicates, and if so, can these predicates naturally appear in many different transaction sets? This is exactly what we have set out to do. The remainder of this paper focuses on what we have found.

4. Translating messages

Part of the challenge facing us as we began this investigation was to choose a set of EDI documents that are in use but not too big. We ended up using EDIFACT messages that Compaq Corporation uses with their business partners. (EDI implementation information and sample documents were downloaded from Compaq's web site at `http://www-.compaq.com/corporate/EDI/` on June 11, 1998.) In this section we examine some of these messages, translate them into our representation, and analyze the results. EDI messages, including EDIFACT messages, do not easily reveal their logical structure, but must be unpacked when they are interpreted by an agent [15]. In the interests of simplicity and focus, the representation we use in our analysis is much closer to the original EDIFACT representation than to Figure 1. We are mapping the EDIFACT definitions to predicates which in turn should be easily mappable into ESΘ format.

4.1. ORDERS document

A sample ORDERS document is shown in Figure 2 and our approximate translation of this document into English is shown in Figure 3. The translation of this document into an ESΘ representation is shown in Appendix A.1. The vocabulary necessary for this document is shown in Table 2 in the appendix.

The EDI message itself is a blanket purchase order from Compaq to one of its suppliers for 10 million pieces (at $50 each) of part 123456. Delivery (by Airborne) is requested by March 1. Payment is due by 45 days from the date of the

1. This is simply a document header, giving primarily addressing information
2. This is message #1 in this document. This is a message of type ORDERS that conforms to standard S, 93A, UN.
3. This is an original blanket order, numbered P1M24987E.
4. The purchase order date is January 1, 1997.
5. This is free text concerning the purchase order itself.
6. This is contract number 123-456.
7. Mistake. See text for explanation.
8. Compaq identifies the seller as the vendor with code `8049P` who is called 'Supplier name'.
9. This provides the address to which the bill should be sent.
10. The buyer is identified as Compaq, specifically the Scotland site (`MFGB`).
11. This provides the address to which the goods should be shipped.
12. Claretta is the purchasing contact.
13. Steve is the sales representative.
14. This is Compaq's tax ID number.
15. U.S. dollars are the currency used in this document.
16. Basic payment terms are that payment is due within 45 days from the date of the order.
17. Discount payment terms are that the discount is given if payment is received by 30 days from the date of the order.
18. The discount given is 2%.
19. Main carriage transport is by Airborne.
20. Transfer of ownership takes place on Compaq's dock.
21. Dispatch conditions are 'collect'.
22. The first line item has Compaq part number `107315-001`.
23. This part number has part number revision level `AA`. The vendor part number is `123456`.
24. The part is described here.
25. 10,000,000 pieces are ordered.
26. Delivery is requested on March 1, 1997.
27. This contains free text related to the first line item.
28. The contract price is 50.
29. This begins the summary of the message.
30. This asserts that there are 29 segments in this message.
31. This asserts that there is 1 message in this document, and this document is referred to as `00...916`.

Figure 3: Approximate translation of ORDERS document

```
 1 UNB+UNOB:1+003897733:01:MFGB-PO
    +PARTNER ID:ZZ+970101:1050
    +00000000000916++ORDERS
 2 UNH+1+ORDERS:S:93A:UN
 3 BGM+221+P1M24987E+9
 4 DTM+4:970101:101
 5 FTX+PUR+3++PURCHASE ORDER BEFORE LINE
    ITEM INSTRUCTIONS
 6 RFF+CT:123-456
 7 RFF+CR:1
 8 NAD+SE+8049P::92++SUPPLIER NAME
 9 NAD+BT+B2::92++COMPAQ COMPUTER
    CORPORATION+P O BOX 692000+HOUSTON+TX
    +77692000+US
10 NAD+BY+B2::92++COMPAQ COMPUTER
    CORPORATION
11 NAD+ST+CM6::92++COMPAQ COMPUTER
    CORPORATION+CCM6 RECEIVING DOCK
    :20555 SH 249+HOUSTON+TX+77070+US
12 CTA+PD+:CLARETTA STRICKLAND-FULTON
13 CTA+SR+:STEVE 10/19/92
14 TAX+9++++++3-00105-5135-3
15 CUX+2:USD:9
16 PAT+1++1:1:D:45
17 PAT+22++1:1:D:30
18 PCD+12:2
19 TDT+20++++:::AIRBORNE
20 LOC+16+COMPAQ DOCK
21 TOD+2+CC+:::ORIGIN COLLECT
22 LIN+000001++107315-001:BP
23 PIA+1+AA:EC+123456:VP
24 IMD+F+8+:::PART DESCRIPTION INFORMATION
25 QTY+21:10000000:PCE
26 DTM+2:970301:101
27 FTX+LIN+3++LINE ITEM COMMENTS
28 PRI+CON:50
29 UNS+S
30 UNT+29+1
31 UNZ+1+000000000000916
```

Figure 2: Sample ORDERS document from Compaq

order and a 2% discount will be given if payment is received within 30 days. Arriving at this translation is not straightforward. In the following we describe some of these problems, including leaving information vital to the message's interpretation out of the message itself (i.e., the message's leanness), idiosyncratic interpretation of data, and the complexity of the message definition. We follow the discussion of the EDI message itself with a discussion of the translation into the ESΘ representation.

In several fields in the ORDERS document, field values are identified as being assigned by one of the parties. The meaning of each of these fields is undefined by Compaq within this document, though it may be defined formally within the interchange agreement or informally by practice. Thus, to send this purchase order to a new trading partner, that trading partner would have to set up its EDI system so that it could properly handle this information that is specific only to the messages from Compaq. And Compaq is not alone here. This is the difficulty of using EDI standards—each company puts information in them that makes them idiosyncratic to their company. Following are some examples from this message in which Compaq follows the standard but still creates a message that is idiosyncratic to Compaq (line numbers refer to Figure 3 and page numbers refer to Compaq's document describing their implementation of the ORDERS message).

According to the EDIFACT standard, Line 1 has a field called "date of preparation." Compaq interprets this field as "date document transmitted." (p. 8) Certainly, these can be different dates. Compaq's interpretation is probably preferred since transmission is more easily verified than preparation. However, Compaq has changed the standard to suit their own needs.

Line 8 asserts that the seller has vendor code `8049P` and that this code is assigned by the buyer (Compaq, in this case). The only way for the seller to know what this means is for the buyer to communicate via some other means (an earlier or later message, a fax, a phone call) the interpretation of this code.

Line 14 states that Compaq's tax ID number is `3-00...`. This also signals that this sale is taxable. If Compaq had used tax ID number `1-76...`, then that would signal that this sale is not taxable. Thus, this piece of data includes both an assertion concerning the appropriate tax ID number and an assertion about the tax status of this potential sale.

Line 23, according to the EDIFACT standard, states that `AA` is the engineering change level; however, Compaq interprets this value as Compaq's part number revision level (p. 30). This is close to, but not the same as, the meaning of "engineering change level."

Line 26, read according to the standard, states that delivery is requested on March 1. Compaq's instructions (p. 33) state that "[f]or blanket purchase orders [of which, this is one], this date will reflect a date in the future. The actual delivery requirements are conveyed in the DELJIT document." Thus, March 1 is *not* the requested delivery date even though that is what the EDI standards says it is. March 1 is certainly "in the future" when compared with January 1, but is this a random date—generated simply to meet the requirements to have a date in this field—or does it have some meaning to either Compaq or the vendor? If it does, it is not stated in

Compaq's instructions. Further, if this were a regular purchase order (i.e., not a blanket order), then is this the date that Compaq wants delivery to begin or to finish? Again, this is not in the instructions so it must be determined either by the interchange agreement or by practice.

Another problem with EDI is that the complexity of the standards makes it difficult to construct messages that conform to them. Line 7 should not be in the document. Compaq's own instructions for the ORDERS document (p. 14) states that this segment is only sent when the last field of the `BGM` segment (see line 3) is `5`. As can be seen, the last field of that segment contains `9`—so this document should *not* contain this segment. The purpose of pointing this out is not to chastise Compaq but to highlight that EDI documents are difficult to create, and that anyone can make a mistake, even a company putting together an instructional handout on how to properly create an EDI document.

Line 8 is defined to have an additional mandatory field after the supplier name. However, Compaq's notes state that this field is to be used only for purchase orders issued by Compaq Brazil (p. 15). Thus, Compaq seems to define two types of mandatory fields: one that is mandatory and one that is mandatory under certain conditions. A similar type of problem crops up with fields that are "conditionally" required. What is it conditional on? It doesn't say in the message definition but knowledge of this condition is vital to putting together a valid message.

Line 9 asserts that this item is being purchased by `B2` (p. 41). This is a mistake. The example for this segment is also in error (p. 8). The instructions say that this value should indicate the routing directions for the return message. Since this message is coming from `MFGB`, this field should contain `MFGB-PO`. This is also an example of an idiosyncratic code that is unique to one of the parties.

Line 10 also contains an error. The line starts `NAD+BY+B2::` while it should, according to Compaq's own instructions, start `NAD+BY+MFGB` (or, e.g., `MFUS`).

Line 14 can include either a U.S. tax ID (taxable), a U.S. tax ID (non-taxable), or a Brazilian tax ID. It is not clear from this document how Compaq can (or, even, *if* it can) signal that the sale is non-taxable if the sale is to a Brazilian company. This is especially vexing considering that the fact that the sale to a U.S. company is taxable (or not) is only *implicitly* asserted in the document.

Line 27 contains free text but Compaq's notes (p. 34) asserts that these are only used by Compaq Houston. First, it does not say how these notes would be used; these are free text notes so they are not going to be interpreted by the computer in any case. The interchange agreement must provide the guidance, specifying how and when this text might be used.

Surprisingly, it does not say anywhere in the message that Compaq is the message's sender. It does state that the interchange sender ID is `003897733`; this may be the universal symbol for Compaq but to verify this fact, access to this "universal" code list, separate from the interchange agreement between trading partners, must be possible.

Similarly, lines 9, 10, & 11 do not state that the entity handling the payment, the entity buying the goods, and the entity receiving the goods are *the same entity*. These entities have the same text string representation but it is not stated that this is the same actual entity.

The above make it fairly clear that the EDI representation is anything but clear. Given this basis, translating the message into an ESΘ representation was sometimes problematic. Consider the following. `Agent(eOrd, cpq), Goal(eBill, cpq), (Goal(eShip, cpq), Agent(ePay, cpq), Goal(eoChg, cpq)`: All these together imply that the agent doing the ordering (`cpq`) is the same as the agent who will receive the bill, the agent who will receive the shipment, the agent who will make the payment, and the agent who will receive ownership of the goods. This information is not in the EDI message itself.

Related to the above, the EDI document contains the text `COMPAQ COMPUTER CORPORATION` three times. The ESΘ representation only contains it once since it represents that the entity being billed, buying, and receiving the shipment are one and the same.

Finally, the ESΘ representation contains the following:

```
& referenceDate(pt,
    ?x:(ordering(eOrd) & Cul(eOrd, x)
         & blanket(eOrd)
         & original(eOrd),
         & terms(eOrd, pt)))
```

This implies that the reference date for payment terms `pt` is `x`, where x is the date that a the original blanket order was placed. The EDI representation asserted that the reference date was the date of the purchase order (p. 24). If the current message were *not* the original order, then it would be possible that the date of this message to differ from the date of the order. The above representation makes this clear.

5. Conclusion

There are two main contributions in this paper. First, we have introduced, discussed and articulated a rather nuanced spanning hypothesis for EDI semantics. The hypothesis, or rather demonstration of its truth, should be seen as essentially a necessary condition for genuinely successful widespread automation in electronic commerce. If the hypothesis is not true in a given domain, how successful can EDI be? So, the hypothesis should be seen as characterizing a programme of research.

Second, we have presented findings from that programme of research. We have analyzed several EDIFACT transaction sets—actually used by the Compaq Corporation—and

discussed one in detail in this work to see whether and how well they fit the spanning hypothesis. Our findings are positive, but qualified. On the positive side, it appeared that the elements used to define the transaction sets we examined can rather naturally fit into the ESΘ semantics developed by Kimbrough, and can do so with considerable economy and prospect for reuse on other transaction sets. We note two qualifications. First, we found what others have found: underspecification of the standards, and idiosyncratic use and interpretation of them. We believe (and surely hope) that a good semantics can contribute materially to overcoming this sort of problem, but this has not yet been demonstrated conclusively. Second, the list of required predicates for the controlled lexicon has grown nontrivially. For example, we started with Table 1 and it has evolved into Table 2 simply by working with these EDI messages. Even so, Table 2 is surely smallish compared to the original EDIFACT definitions.

Much is left on our plate. We must translate more EDIFACT documents into our intermediate representation (as shown in the appendix) to investigate the vocabulary that might be needed. We also must translate this intermediate representation into an ESΘ representation to discover the logical structures that are needed. Given these logical structures we will be in a better position to investigate how complex systems would have to be to make sense of this representation. This is a vital step for we believe that the ESΘ representation would make a powerful and useful ACL. The ESΘ representation should have other uses as well. We can envision it serving as a meta-language for formally representing the meaning of messages. It also might serve as the target language into which an EDI message, from whatever EDI standard, is unpacked once it is received.

Much remains to be discovered, but there are reasons for optimism.

A. Sample messages

A.1. ORDERS message

```
ordering(eOrd)
 & Agent(eOrd, cpq)
 & Addressee(eOrd, vend)
 & Theme(eOrd, [item1])
 & Cul(eOrd, '970101')
 & Type(eOrd, 'ORDERS')
 & blanket(eOrd)
 & original(eOrd)
 & referenceNumber(eOrd, 'P1M24987E')
 & freeText(eOrd, 'PURCHASE ORDER
  BEFORE LINE ITEM INSTRUCTIONS')
 & contract(eOrd, ct)
 & referenceNumber(ct, '123-456')
 & referenceNumber(vend, '8049P')
 & name(vend, 'SUPPLIER NAME')
 & billing(eBill)
 & Agent(eBill, vend)
 & Goal(eBill, cpq)
 & Location(eBill, bt)
 & name(cpq, 'COMPAQ COMPUTER CORPORATION')
 & addressLine(bt, 1, 'PO BOX 692000')
 & addressCity(bt, 'HOUSTON')
 & addressSubCountry(bt, 'TX')
 & addressPostalID(bt, '77692000')
 & addressCountry(bbt, 'US')
 & referenceNumber(cpq, 'MFGB')
 & shipping(eShip)
 & Agent(eShip, vend)
 & Goal(eShip, cpq)
 & Location(eShip, st)
 & addressLine(st, 1, 'CCM6 RECEIV-
ING DOCK')
 & addressLine(st, 2, '20555 SH 249')
 & addressCity(st, 'HOUSTON')
 & addressSubCountry(st, 'TX')
 & addressPostalID(st, '77070')
 & addressCountry(st, 'US')
 & method(eShip, 'MAIN CARRIAGE TRANSPORT')
 & carrier(eShip, 'AIRBORNE')
 & contact(eOrd, cpq,
     'CLARETTA STRICKLAND-FULTON')
 & contact(eOrd, vend, 'STEVE 10/19/92')
 & taxIDNumber(cpq, '3-00105-5135-3')
 & Type('3-00105-5135-3', 'US')
 & Type('3-00105-5135-3', 'TAXABLE')
 & currency(eOrd, 'USD')
 & paying(ePay)
 & Agent(ePay, cpq)
 & Goal(ePay, vend)
 & Location(ePay, 'ORIGIN COLLECT')
 & terms(ePay, [pt, disc])
 & Type(pt, 'BASIC')
 & referenceDate(pt,
     ?x:(ordering(eOrd) & Cul(eOrd, x)
          & Theme(eOrd, eOrd)
          & blanket(eOrd)
          & original(eOrd)
          & paymentTerms(eOrd, pt)))
 & units(pt, 'DAY')
 & quantity(pt, 45)
 & Type(disc, 'DISCOUNT')
 & referenceDate(disc,
     ?x:(ordering(eOrd) & Cul(eOrd, x)
          & Theme(eOrd, eOrd)
          & blanket(eOrd)
          & original(eOrd)
          & paymentTerms(eOrd, disc)))
 & units(disc, 'DAY')
 & quantity(disc, 30)
 & units(disc, 'PCT')
 & quantity(disc, 2)
 & ownershipChanging(eOChg)
 & Agent(eOChg, vend)
 & Goal(eOChg, cpq)
```

```
& Location(eOChg, 'COMPAQ DOCK')
& itemNumber(item1, '130918-001', c1)
& catalog(c1, cpq)
& itemNumber(item1, '123456', c2)
& catalog(c2, vend)
& Type(c2, 'AA')
& Type(item1, 'PRODUCT')
& description(item1, 'PART
    DESCRIPTION INFORMATION')
& quantity(item1, 10000000)
& units(item1, 'PCE')
& price(item1, 50)
```

References

[1] M. Covington, "Toward a new type of language for electronic commerce," in *Proceedings of the Twenty-Ninth Annual Hawaii International Conference on System Sciences, Vol. IV, Information Systems—Organizational Systems and Technology* (J. F. Nunamaker, Jr. and R. H. Sprague, Jr., eds.), (Los Alamitos, CA), pp. 329–336, IEEE Press, 1996.

[2] R. M. Lee, "CANDID: A formal language for electronic contracting," tech. rep., EURIDIS: Erasmus University Research Institute for Decision and Information Systems, Erasmus University, Rotterdam, The Netherlands, 29 August 1996.

[3] R. M. Lee, *CANDID—A Logical Calculus for Describing Financial Contracts*. PhD thesis, University of Pennsylvania, Department of Decision Sciences (now Department of Operations and Information Management), 1980. Working paper 80-06-2.

[4] F. Lehmann, "Machine-negotiated, ontology-based EDI (electronic data interchange)," in *Electronic Commerce: Current Research Issues and Applications* (N. R. Adam and Y. Yesha, eds.), vol. 1028 of *Lecture Notes in Computer Science*, pp. 27–46, Berlin, Germany: Springer, 1996. ISBN: 3-540-60738-2.

[5] XML/EDI Group, The, "Home page for the XML/EDI Group." Web page, January 1998. `http://www.xmledi.net`.

[6] S. O. Kimbrough and S. A. Moore, "On obligation, time, and defeasibility in systems for electronic commerce," in *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences, Volume III, Information Systems: DSS/Knowledge-Based Systems* (J. F. Nunamaker, Jr. and R. H. Sprague, Jr., eds.), (Los Alamitos, CA), pp. 493–502, IEEE Computer Society Press, 1993. File: HICSS-93-final-1.2.

[7] I. A. Smith and P. R. Cohen, "Toward a semantics for an agent communications language based on speech-acts," in *Proceedings IAAI*, pp. 24–31, 1996.

[8] S. O. Kimbrough, "On representation schemes for promising electronically," *Decision Support Systems*, vol. 6, no. 2, pp. 99–122, 1990.

[9] S. O. Kimbrough, "On electronic commerce, subatomic semantics and Cæsar's stabbing," in *Proceedings of the Thirtieth Hawaii International Conference on System Sciences* (R. H. Sprague, Jr., ed.), (Los Alamitos, CA), IEEE Press, 1997. File: hicss97-sok-subatomic-final.

[10] S. O. Kimbrough, "On ESΘ theory and the logic of the X12 date/time qualifiers," in *Proceedings of the Thirty-First Hawai'i International Conference on System Sciences* (R. H. Sprague, Jr., ed.), (Los Alamitos, CA), IEEE Press, 1998. File: flbcx12.tex.

[11] S. O. Kimbrough, "Sketch of a basic theory for a formal language for business communication," in *Proceedings of the Thirty-First Hawai'i International Conference on System Sciences* (R. H. Sprague, Jr., ed.), (Los Alamitos, CA), IEEE Press, 1998. File: flbch98.tex.

[12] S. O. Kimbrough and S. A. Moore, "On automated message processing in electronic commerce and work support systems: Speech act theory and expressive felicity," *ACM Transactions on Information Systems*, vol. 15, no. 4, pp. 321–367, October 1997.

[13] T. Parsons, *Events in the Semantics of English: A Study in Subatomic Semantics*. Current Studies in Linguistics, Cambridge, MA: The MIT Press, 1990. ISBN: 0-262-66093-8.

[14] R. Larson and G. Segal, *Knowledge of Meaning: An Introduction to Semantic Theory*. Cambridge, Massachusetts: The MIT Press, 1995. ISBN: 0-262-62100-2.

[15] S. O. Kimbrough and Y.-H. Tan, "On lean messaging with wrapping and unfolding for electronic commerce." Manuscript, June 1998. File: lean-sok-tan-19980324.TEX.

[16] S. A. Moore, "Categorizing automated messages," *Decision Support Systems*, forthcoming.

Standard roles	
Addressee(x, a)	a is the addressee of x
Agent(e, x)	the agent of event e is x
Cul(e, d)	e culminates on date d
Location(e, x)	e takes place at location x
Theme(e, x)	x is the theme of e
Type(a, b)	a can be subtyped as b
Verbs	
billing(x)	x is a billing event
confirming(x)	x is a confirming event
forecasting(x)	x is a forecasting event
ordering(x)	x is an ordering event
ownershipChanging(x)	x is an ownership changing event
paying(x)	x is a paying event
shipping(x)	x is a shipping event
Miscellaneous	
addressCity(x, y)	the city of x is y
addressCountry(x, y)	the country of x is y
addressLine(x, y, z)	address line number y for x is z
addressPostalID(x, y)	the postal id for x is y (ZIP code for US)
addressSubCountry(x, y)	the sub-country identifier for x is y (state for US)
billTo(x, y)	bill to y for x
blanket(x)	x is a blanket action or event
carrier(x, y)	the carrier for delivery of x is y
catalog(x, y)	x is a catalog of y
contact(x, y, z)	the contact for x at party y is z
contract(x, y)	the contract number for x is y
currency(x, y)	the currency for x is expressed in y
description(x, y)	x is described by y
freeText(x, y)	free text associated with x is y
itemNumber(x, y, z)	item x has part number y in catalog z
method(x, y)	the method for x is y
name(x, y)	the name of x is y
original(x)	x is an original message
price(x, y)	the price of y is x
quantity(x, y)	there are y units of x
referenceNumber(x, y)	the reference number for x is y
schedule(x, y)	the delivery schedule for x is y
seller(e, x)	x is the seller in the e event
taxIDNumber(x, y)	the tax id number of x is y
terms(e, x)	the terms for event e is x
units(x, y)	the units of x are expressed in y's

Table 2: Application vocabulary