Balancing Expressiveness and Simplicity in an Interlingua for Task Based Dialogue

Lori Levin, Donna Gates, Dorcas Wallace, Kay Peterson, Alon Lavie
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213
email: lsl@cs.cmu.edu

Fabio Pianesi, Emanuele Pianta, Roldano Cattoni, Nadia Mana
IRST-itc, Italy

Abstract

In this paper we compare two interlingua representations for speech translation. The basis of this paper is a distributional analysis of the C-STAR II and NESPOLE databases tagged with interlingua representations. The C-STAR II database has been partially re-tagged with the NESPOLE interlingua, which enables us to make comparisons on the same data with two types of interlinguas and on two types of data (C-STAR II and NESPOLE) with the same interlingua. The distributional information presented in this paper show that the NESPOLE interlingua maintains the language-independence and simplicity of the C-STAR II speech-act-based approach, while increasing semantic expressiveness and scalability.

1 Introduction

Several speech translation projects have chosen interlingua-based approaches because of its convenience (especially in adding new languages) in multi-lingual projects. However, interlingua design is notoriously difficult and inexact. The main challenge is deciding on the grain size of meaning to represent and what facets of meaning to include. This may depend on the domain and the contexts in which the translation system is used. For projects that take place at multiple research sites, another factor becomes important in interlingua design: if the interlingua is too complex, it cannot be used reliably by researchers at remote sites. Furthermore, the interlingua should not be biased toward one family of languages. Finally, an interlingua should clearly distinguish general and domain specific components for easy scalability and portability between domains.

Sections 2 and 3 describe how we balanced the factors of grain-size, language independence, and simplicity in two interlinguas for speech translation projects — the C-STAR II Interchange Format (Levin et al., 1998) and the NESPOLE Interchange Format. Both interlinguas are based in the framework of *domain actions* as described in (Levin et al., 1998). We will show that the NESPOLE interlingua has a finer grain-size of meaning, but is still simple enough for collaboration across multiple research sites, and still maintains language-independence.

Section 4 will address the issue of scalability of interlinguas based on domain actions to larger domains. The basis of Section 4 is a distributional analysis of the C-STAR II and NESPOLE databases tagged with interlingua representations. The C-STAR II database has been partially re-tagged with the NESPOLE interlingua, which enables us to make comparisons on the same data with two types of interlinguas and on two types of data (C-STAR II and NESPOLE) with the same type of interlingua.

2 The C-STAR II Domain, Database, and Interlingua

The C-STAR II interlingua (Levin et al., 1998) was developed between 1997 and 1999 for use in the C-STAR II 1999 demo (www.c-star.org).

```
c: can I have some flight times
   that would leave some time around June sixth
a: the there are several flights leaving D C
   there'd  be one  at one twenty four
   there's a three fifty nine flight
   that arrives at four fifty eight
   ...
   what time would you like to go
c: I would take the last one that you mentioned
   ...
a: what credit card number would you like
   to reserve this with
c: I  have a visa card
   and  the number is double oh five three
   three one one six
   ninety nine eighty seven
a  okay
c: the expiration date is eleven ninety seven
...
a  okay they should be ready tomorrow
c: okay thank you very much
```

Figure 1: **Excerpt from a** C-STAR **II dialogue**

with six participating research sites. The semantic domain was travel, including reservations and payments for hotels, tours, and transportation. Figure 1 shows a sample dialogue from the C-STAR II database. (`C` is the client and `a` is the travel agent.) The C-STAR II database contains 2278 English sentences and 7148 non-English (Japanese, Italian, Korean) sentences tagged with interlingua representations. Most of the database consists of transcripts of role-playing conversations.

The driving concept behind the C-STAR II interlingua is that there are a limited number of actions in the domain — requesting the price of a room, telling the price of a room, requesting the time of a flight, giving a credit card number, etc. — and that each utterance can be classified as an instance of one of these *domain actions*. Figure 2 illustrates the components of the C-STAR II interlingua: (1) the speaker tag, in this case `c` for client, (2) a speech act (`request-action`), (3) a list of concepts (`reservation, temporal, hotel`), (4) arguments (e.g., `time`), and (5) values of arguments. The C-STAR II interlingua specification document contains definitions for 44 speech acts, 93 concepts, and 117 argument names. The *domain action* is the part of the interlingua consisting of the speech act and concepts, in this case `request-action+reservation+temporal+hotel`. The domain action does not include the list of argument-value pairs.

First it is important to point out that domain actions are created compositionally. A domain action consists of a speech act followed by zero or more concepts. (Recall that argument-value pairs are not part of the domain action.) The NESPOLE interlingua includes 65 speech acts and 110 concepts. An interlingua specification document defines the legal combinations of speech acts and arguments.

The linguistic justification for an interlingua based on domain-actions is that many travel domain utterances contain fixed, formulaic phrases (e.g., *can you tell me; I was wondering; how about; would you mind, etc.*) that signal domain actions, but either do not translate literally into other languages or have a meaning that is sufficiently indirect that the literal meaning is irrelevant for translation. To take two examples, *how about* as a signal of a suggestion does not translate into other languages with the words corresponding to *how* and *about*. Also, *would you mind* might translate literally into some European languages as a way of signaling a request, but the literal meaning of *minding* is not relevant to the translation, only the fact that it signals politeness.

The measure of success for the domain-action based interlingua (as described in (Levin et al., 2000a)) is that (1) it covers the data in the C-STAR II database with less than 8% no-tag rate, (2) inter-coder agreement across research sites is reasonably high: 82% for speech acts, 88% for concepts, and 65% for domain actions, and (3) end-to-end translation results using an analyzer and generator written at different sites were about the same as end-to-end translation results using an analyzer and generator written at the same site.

3 The NESPOLE Domain, Database, and Interlingua

The NESPOLE interlingua has been under development for the last two years as part of the NESPOLE project (`http://nespole.itc.it`). Fig-

```
I would like to make a hotel reservation for the fourth through
the seventh of july

c:request-action+reservation+temporal+hotel
(time=(start-time=md4, end-time=(md7, july)))
```

Figure 2: **Example of a** C-STAR **II interlingua representation**

ure 3 shows a NESPOLE dialogue. The NE-SPOLE domain does not include reservations and payments, but includes more detailed inquiries about hotels and facilities for ski vacations and summer vacations in Val di Fiemme, Italy. (The tourism board of the Trentino area is a partner of the NESPOLE project.) Most of the database consists of transcripts of dialogues between an Italian-speaking travel agent and an English or German speaker playing the role of a traveller.

There are fewer fixed, formulaic phrases in the NESPOLE domain, prompting us to move toward domain actions that are more general, and also requiring more detailed interlingua representations. Changes from the C-STAR II interlingua fall into several categories:

1. **Extending semantic expressivity and syntactic coverage:** Increased coverage of modality, tense, aspect, articles, fragments, coordinate structures, number, and rhetorical relations. In addition, we have added more explicit representation of grammatical relations and improved capabilities for representing modification and embedding.

2. **Additional Domain-Specific Concepts:** New concepts include giving directions, describing sizes and dimensions of objects, traveling routes, equipment and gear, airports, tourist services, facilities, vehicles, information objects (brochures, web pages, rules and regulations), hours of operation of businesses and attractions, etc.

3. **Utterances that accompany multimodal gestures:** The NESPOLE system includes capabilities to share web pages and draw marks such as circles and arrows on web pages. The interlingua was extended to cover colord, descriptions of two-dimensional objects, and actions of showing.

4. **General concept names from WordNet:** The NESPOLE interlingua includes conventions for making new concept names based on WordNet synsets.

5. **More general domain actions replacing specific ones:** For example, replacing `hotel` with `accommodation`.

Interlinguas based on domain actions contrast with interlinguas based on lexical semantics (Dorr, 1993; Lee et al., 2001; Goodman and Nirenburg, 1991). A lexical-semantic interlingua includes a representation of predicates and their arguments. For example, the sentence *I want to take a vacation* has a predicate *want* with two arguments *I* and *to take a vacation*, which in turn has a predicate *take* and two arguments, *I* and *a vacation*. Of course, predicates like *take* may be represented as word senses that are less language-dependent like `participate-in`. The strength and weakness of the lexical-semantic approach is that it is less domain dependent than the domain-action approach.

In order to cover the less formulaic utterances of the NESPOLE domain, we have taken a step closer to the lexical-semantic approach. However, we have maintained the overall framework of the domain-action approach because there are still many formulaic utterances that are better represented in a non-literal way. Also, in order to abstract away from English syntax, concepts such as disposition, eventuality, and obligation are not represented in the interlingua as argument-taking main verbs in order to accommodate languages in which these meanings are

```
c: and I have some questions about coming about a trip I'm gonna be taking to Trento
a: okay what are your questions
c: I currently have a hotel booking at the
   Panorama-Hotel in Panchia but at the moment I have no idea how to get to my hotel from Trento
   and I wanted to ask what would be the best way for me to get there
a: okay I'm gonna show you a map that and then describe the directions to you
   okay so right so you will arrive in the train station in Trento
   the that is shown in the middle of the map stazione FFSS
   and just below that here is a bus stop labeled number forty
   so  okay  on the map that I'm showing you here
   the hotel is the orange building off on the right hand side
...
c: I also wanted to ask about skiing in the area once I'm in Panchia
a: all right  just a moment and I'll show you another map
c: okay
a: okay  so  on the map you see now Panchia is right in the center of the map
c: I see it
```

Figure 3: **Excerpt from a** Nespole **dialogue**

represented as adverbs or suffixes on verbs. Figure 4 shows the Nespole interlingua representation corresponding to the C-star II interlingua in Figure 2. The specification document for the Nespole interlingua defines 65 speech acts, 110 concepts, 292 arguments, and 7827 values grouped into 222 value classes. As in the C-star II interlingua, domain actions are defined compositionally from speech acts and arguments in combinations that are allowed by the interlingua specification.

3.1 Comparison of Nespole and C-star II Interlinguas

It is useful to compare the Nespole and C-star II Interlinguas in expressivity, language independence, and simplicity.

Expressivity of the Nespole **interlingua, Argument 1:** The metric we use for expressivity is the no-tag rate in the databases. The no-tag rate is the percentage of sentences that cannot be assigned an interlingua representation by a human expert. The C-star II database tagged with C-star II interlingua had a no-tag rate of 7.3% (Levin et al., 2000a). The C-star II database tagged with Nespole interlingua has a no-tag rate of 2.4%. More than 300 English sentences in the C-star II database that were not covered by the C-star II interlingua are now covered by the Nespole interlingua. (See Table 2.) We conclude from this that

the Nespole interlingua is more expressive in that it covers more data.

Language-independence of the Nespole **interlingua:** We do not have a numerical measure of language-independence, but we note that interlinguas based on domain actions are particularly suitable for avoiding translation mismatches (Dorr, 1994), particularly head-switching mismatches (e.g., *I just arrived* and *Je vient d'arriver* where the meaning of recent past is expressed by an adverb *just* or a syntactic verb *vient (venir)*.) Interlinguas based on domain actions resolve head-switching mismatches by identifying the types of meanings that are often involved in mismatches — modality, evidentiality, disposition, and so on — and assigning them a representation that abstracts away from predicate argument structure. Interlinguas based on domain actions also neutralize the different ways of expressing indirect speech acts within and across languages (for example, *Would you mind...*, *I was wondering if you could....*, and *Please....* as ways of requesting an action). Although Nespole domain actions are more general than C-star II domain actions, they maintain language independence by abstracting away from predicate-argument structure.

Simplicity and cross-site reliability of the Nespole **interlingua:** Simplicity of an interlingua is measured by cross-site reliability in

```
I would like to make a hotel reservation for the fourth through
the seventh of july

C-star II Interlingua:
c:request-action+reservation+temporal+hotel
(time=(start-time=md4, end-time=(md7, july)))

Nespole Interlingua:
c:give-information+disposition+reservation+accommodation
  (disposition=(who=i, desire),
   reservation-spec=(reservation, identifiability=no),
   accommodation-spec=hotel,
   object-time=(start-time=(md=4), end-time=(md=7, month=7, incl-excl=inclusive)))}
```

Figure 4: **Example of** NESPOLE **interlingua representation**

inter-coder agreement and end-to-end translation performance. At the time of writing this paper we have not conducted cross-site inter-coder agreement experiments using the NESPOLE interlingua. We have, however, conducted cross-site evaluations (Lavie et al., 2002), in which the analyzer and generator were written at different sites. Experiments at the end of C-STAR II showed that cross-site evaluations were comparable to intra-site evaluations (analyzer and generator written at the same site) (Levin et al., 2000b). NESPOLE evaluations so far show a loss of cross-site reliability: intra-site evaluations are noticeably better than cross-site evaluations, as reported in (Lavie et al., 2002). This seems to indicate that developers at different sites have a lower level of agreement on the NESPOLE interlingua. However there are other possible explanations for the discrepancy — for example developers at different sites may have focused their development on different sub-domains — that are currently under investigation.

4 Scalability of the NESPOLE Interlingua

The rest of this paper addresses the scalability of the NESPOLE interlingua. A possible criticism of domain actions is that they are domain dependent and that the number of domain actions might increase too quickly with the size of the domain. In this section, we will examine the rate of increase in the number of domain actions as a function of the amount of data and the diversity of the data.

Differences in the C-STAR **and** NESPOLE **Domains:** We will first show that the C-STAR and NESPOLE domains are significantly different even though they both pertain to travel. The combination of the two domains is therefore significantly larger than either domain alone.

In order to demonstrate the differences between the C-STAR travel domain and the NESPOLE travel domain, we measured the overlap in vocabulary. The numbers in Table 4 are based on the first 7900 word tokens in the C-STAR English database and the first 7900 word tokens in the NESPOLE English database. The table shows the number of unique word types in each database, the number of word types that occur in both databases, and the number of word types that occur in one of the databases, but not in the other. In each database, about half of the word types overlap with the other database. The non-overlapping vocabulary (402 C-STAR word types and 344 NESPOLE word types) indicates that the two databases cover quite different aspects of the travel domain.

Scalability: Argument 1: We will now begin to address the issue of scalability of the domain action approach to interlingua design. Our first argument concerns the number of

	Number of unique word types
CSTAR English	745
Nespole English	687
Word types in both CSTAR and Nespole	343
Words types in CSTAR not in Nespole	402
Words types n Nespole not in CSTAR	344

Table 1: **Number of overlapping word types in the** C-STAR **English and** NESPOLE **English databases**

	SA	Con.	Snts.	Domain Actions
Old C-STAR **English**	44	93	2278	358
New C-STAR **English**	65	110	2564	452
NESPOLE **English**	65	110	1446	337
NESPOLE **German**	65	110	3298	427
NESPOLE **Italian**	65	110	1063	206

Table 2: **Number of unique domain actions in interlingua databases**

speech acts and concepts in the combined C-STAR/NESPOLE domain. The C-STAR II interlingua, designed for coverage of the C-STAR travel domain, included 44 speech acts and 93 concepts. The NESPOLE interlingua, designed for coverage of the combined C-STAR and NESPOLE domains, has 65 speech acts and 110 concepts. Thus a relatively small increase in the number of speech acts and concepts is required to cover a significantly larger domain.

The increased size of the C-STAR/NEPSOLE domain is reflected in the number of arguments and values. The C-STAR II interlingua contained definitions for 117 arguments, whereas the NESPOLE interlingua contains definitions for 292 arguments. The number of values for arguments also has increased significantly in the NESPOLE domain. There are 7827 values grouped into 222 classes (airport names, days of the week, etc.).

Distributional Data: number of domain actions in each database: Next we will present distributional data concerning the number of domain actions as a function of database size. We will compare several databases: Old C-STAR English (around 2278 sentences tagged with C-STAR II interlingua), New C-STAR English (2564 sentences tagged with NESPOLE interlingua, including the 2278 sentences from Old

C-STAR English), NESPOLE English, NESPOLE German, and NESPOLE Italian. Table 2 shows the number of sentences and the number of domain actions in each database. The number of domain actions refers to the number of types, not tokens, of domain actions.

Distributional data: Coverage of the top 50 domain actions: Table 3 shows the percentage of each database that is covered by the 5, 10, 20, and 50 most frequent domain actions in that database. For each database, the domain actions were ordered by frequency. The percentage of sentences covered by the top-n domain actions was then calculated. For this experiment, we separated sentences spoken by the traveller (client) and sentences spoken by the travel agent (agent). C-STAR data in Table 3 refers to 2564 English sentences from the C-STAR database that were tagged with NESPOLE interlingua. NESPOLE data refers to the English portion of the NESPOLE database (1446 sentences). Combined data refers to the combination of the two (4014 sentences).

Two points are worth noting about Table 3. First, the NESPOLE agent data has a higher coverage rate than the NESPOLE client data. That is, more data is covered by the top-n domain actions. This may be because there was was

Domain Actions	Top 5	Top 10	Top 20	Top 50
Client				
C-star **data**	33.6	42.7	53.1	66.7
Nespole **data**	31.7	43.5	53.9	66.5
Combined data	31.6	40.0	50.3	62.9
Agent				
C-star **data**	33.8	42.8	54.1	67.3
Nespole **data**	39.0	47.8	56.1	71.4
Combined data	33.6	41.5	51.7	64.0

Table 3: **DA Coverage using** Nespole **interlingua on English data for both** C-star **and** Nespole

only a small amount of English agent data and it was spoken by non-native speakers. Second, the combined data has a slightly lower coverage rate than either the C-star or Nespole databases alone. This is expected because, as shown above, the combined domain is significantly more diverse than either domain by itself.

Scalability: Argument 2: Table 3 provides additional evidence for the scalability of the Nespole interlingua to larger domains. In the combined C-star and Nespole domain, the top 50 domain actions cover only slightly less data than the top 50 domain actions in either domain separately. There is not, in fact, an explosion of domain actions when the two C-star and Nespole domains are combined.

Distributional Data: domain actions as a function of database size: Table 3 shows that in each of our databases, the 50 most frequent domain actions cover approximately 65% of the sentences. The next issue we address is the nature of the "tail" of less frequent domain actions covering the remainder of the data.

Figure 5 shows the number of domain actions as a function of data set size. Sampling was done for intervals of 25 sentences starting at 100 sentences. For each sample size **s** there was ten-fold cross-validation. Ten random samples of size **s** were chosen, and the number of different domain actions in each sample was counted. The average of the number of domain actions in each of the ten samples of size **s** are plotted in Figure 5. The four databases represented in Figure 5 are
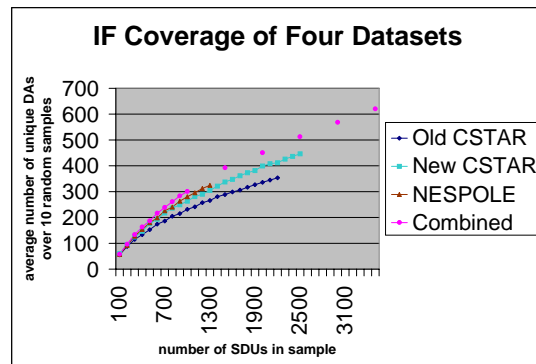

Figure 5: *Number of domain actions as a function of database size*

the C-star English database tagged with C-star II interlingua, the C-star II database tagged with Nespole interlingua, the Nespole English database, and the combined C-star and Nespole English databases.

Expressivity, Argument 2: Figure 5 provides evidence for the increased expressivity of the Nespole interlingua. In contrast to Table 3, which deals with samples containing the most frequent domain actions, the samples plotted in Figure 5 contain random mixtures of frequent and non-frequent domain actions. The curve representing the C-star data with C-star II interlingua is the slowest growing of the four curves. This is because the grain-size of meaning represented in the C-star II interlingua was larger than in the Nespole interlin-

gua. Also many infrequent domain actions were not covered by the C-star II interlingua. The faster growth of the curve representing the C-star data with Nespole interlingua indicates improved expressivity of the Nespole interlingua — it covers more of the infrequent domain actions. The highest curve in Figure 5 represents the combined C-star and Nespole domains. This curve is higher than the others because, as shown above, the two travel domains are significantly different from each other.

Expressivity and Simplicity, the right balance: Comparing Table 3 and Figure 5, we argue that the Nespole interlingua strikes a good balance between expressivity and simplicity. Table 3 shows evidence for the simplicity of the Nespole interlingua: Only 50 domain actions are needed to cover 60-70% of the sentences in the database. Figure 5 shows evidence for expressivity: because domain actions are compositionally formed from speech acts and concepts, it is possible to form a large number of low-frequency domain actions in order to cover the domain. Over 600 domain actions are used in the combined C-star and Nespole domains.

5 Conclusions

We have presented a comparison of a purely domain-action-based interlingua (the C-star II interlingua) and a more expressive, but still domain-action-based interlingua (the Nespole interlingua). The data that we have presented show that the more expressive interlingua has better coverage of the domain (a decrease from 7.3% to 2.4% uncovered data in the C-star II domain) and can also scale up to larger domains without an explosion of domain actions. Thus we have a reasonable compromise between simplicity and expressiveness of the interlingua.

Acknowledgments

References

Bonnie J. Dorr. 1993. *Machine Translation: A View from the Lexicon*. The MIT Press, Cambridge, Massachusetts.

Bonnie J. Dorr. 1994. Machine Translation Divergences: A Formal Description and Proposed Solution. *Computational Linguistics*, 20(4):597–633.

Kenneth Goodman and Sergei Nirenburg. 1991. *The KBMT Project: A Case Study in Knowledge-Based Machine Translation*. Morgan Kaufmann, San Mateo, CA.

Alon Lavie, Florian Metze, Roldano Cattoni, and Erica Constantini. 2002. A Multi-Perspective Evaluation of the NESPOLE! Speech-to-Speech Translation System. In *Proceedings of Speech-to-Speech Translation: Algorithms and Systems*.

Young-Suk Lee, W. Yi, Clifford Weinstein, and Stephanie Seneff. 2001. Interlingua-based broad-coverage korean-to-english translation. In *Proceedings of HLT*, San Diego.

Lori Levin, Donna Gates, Alon Lavie, and Alex Waibel. 1998. An Interlingua Based on Domain Actions for Machine Translation of Task-Oriented Dialogues. In *Proceedings of the International Conference on Spoken Language Processing (IC-SLP'98)*, pages Vol. 4, 1155–1158, Sydney, Australia.

Lori Levin, Donna Gates, Alon Lavie, Fabio Pianesi, Dorcas Wallace, Taro Watanabe, and Monika Woszczyna. 2000a. Evaluation of a Practical Interlingua for Task-Oriented Dialogue. In *Workshop on Applied Interlinguas: Practical Applications of Interlingual Approaches to NLP*, Seattle.

Lori Levin, Alon Lavie, Monika Woszczyna, Donna Gates, Marsal Gavaldà, Detlef Koll, and Alex Waibel. 2000b. The Janus-III Translation System. *Machine Translation*.