

Human-Machine Conversations to Support Multi-Agency Missions*

Alun Preece^a Dave Braines^b Diego Pizzocaro^a Christos Parizas^a
A.D.Preece@cs.cf.ac.uk dave_braines@uk.ibm.com PizzocaroD@cardiff.ac.uk ParizasC@cardiff.ac.uk

^aSchool of Computer Science and Informatics, Cardiff University, Cardiff, UK

^bEmerging Technology Services, IBM United Kingdom Ltd, Winchester, UK

In domains such as emergency response, environmental monitoring, policing and security, sensor and information networks are deployed to assist human users across multiple agencies to conduct missions at or near the “front line”. These domains present challenging problems in terms of human-machine collaboration: human users need to task the network to help them achieve mission objectives, while humans (sometimes the same individuals) are also sources of mission-critical information. We propose a natural language-based conversational approach to supporting human-machine working in mission-oriented sensor networks. We present a model for human-machine and machine-machine interactions in a realistic mission context, and evaluate the model using an existing surveillance mission scenario. The model supports the flow of conversations from full natural language to a form of Controlled Natural Language (CNL) amenable to machine processing and automated reasoning, including high-level information fusion tasks. We introduce a mechanism for presenting the gist of verbose CNL expressions in a more convenient form for human users. We show how the conversational interactions supported by the model include requests for expansions and explanations of machine-processed information.

I. Introduction

A mission-oriented sensor network (MOSN) must support high-level tasking of network resources in terms of mission objectives, and enable exploitation of soft (human) sources in addition to physical sensing assets. These requirements involve human-machine interaction: users need to be able to request information from the network, while also being sources of information. MOSNs are increasingly deployed in domains such as emergency response, environmental monitoring, policing and security, to assist human users — typically across a number of partner agencies — to conduct missions at or near the “front line”. MOSNs have the potential to empower individuals in the field who, prior to the widespread provision of mobile information and communication platforms, have not traditionally been able to benefit from the best-available actionable information [2]. MOSN technology is becoming increasingly service-oriented, offering a range of capabilities from the identification of relevant sources, to the automatic generation of queries and sensor tasking requests, to the composition and invocation of useful information-processing services, to the selection of appropriate dissemination mechanisms which take into account the capabilities of an end-user’s (mobile) device. Many of the technical elements required for MOSNs are discussed in [8].

In this paper we address the need for human-machine interaction in MOSNs by proposing a natural language-based *conversational* approach aimed at making it easier and more convenient for users in the field to access mission-supporting services. We introduce a model for

human-machine and machine-machine interactions that includes support for: (1) requests for information, (2) provision of information, and (3) human-machine reasoning and information fusion. The approach is underpinned by the use of *controlled natural language* (CNL) to provide an information representation that is easily machine processable (with low complexity and no ambiguity) while also being human-readable [11]. A CNL is a subset of a natural language (NL), commonly English, with restricted syntax and vocabulary. For our purposes, using a CNL facilitates clearer communication between human and system, and also enables the system to act directly on the information without the need to transform to/from another technical representation, supporting human-machine reasoning and information fusion [10] in the MOSN context. Several controlled natural languages exist; we selected a form of Controlled English known as ITA Controlled English (CE) [5] for compatibility with related research efforts. A brief guide to CE syntax and modelling is given in the appendix. An example statement in CE syntax is shown below; this identifies an individual known to be a high-value target (HVT):

```
there is a person named p670467 that
    is known as ‘John Smith’ and
    is a high value target.
```

While it is possible for (trained) humans to communicate directly in CNL, for convenience we aim to enable conversations that flow from natural language to CNL and back again, through an exchange of messages we call *cards*. Section II summarises the kinds of interactions we aim to support, with examples. Section III describes our conversation model in terms of the primitive kinds of interaction and valid sequences. Section IV demonstrates how the model

*This work is based on an earlier work: “Human-Machine Conversations to Support Mission-Oriented Information Provision” in MiSeNet’ 13, ©ACM 2013. <http://dx.doi.org/10.1145/2509338.2509342>.

can be used to support realistic exchanges in a MOSN context, using a scenario from previously-published work. Section V describes a prototype implementation of the conversational protocol, while Section VI provides evaluation and discussion, including some refinements to the protocol following experimentation with the prototype.

II. Human-machine conversations

We focus on supporting three main kinds of interaction:¹

human→machine interactions where the purpose of the interaction is to mediate between NL and CE forms of human-provided content. The human initiates an interaction in NL and the machine feeds back CE, prompting the human to refine the CE and agree an unambiguous CE form of the content. Example: an officer on patrol reports a suspicious vehicle at a location by means of a text message from their mobile device; the software agent on their device asks them to confirm their message in CE format (“Did you mean...?”). Note that the human’s content could be a question or statements, and the confirmed form will correspondingly be a CE query (“is it true that the vehicle X is a threat?”) or facts (“the vehicle X is a threat”).

machine→human interactions where the purpose of the interaction is to inform a human or ask them for information. While it is possible to use CE for this purpose, it is often more convenient to present the gist of full CE in a more compact form, for which templates can be used. Example: the information broker agent sends a brief “gist” report to a human analyst indicating the vehicle is associated with a known high-value target. Commonly, a human receiving a gist report will ask for it to be expanded so they can see the full (CE) information behind it; they may also wish to obtain explanations (CE rationale) for some or all of that information. In addition to CE content, communications may have other kinds of linked content, for example imagery or a reference to a document.

machine→machine interactions where the purpose of the interaction is to exchange information between software agents. The conversation is carried out through an exchange of CE content. Example: the CE from the patrol officer in the above example is communicated to an information broker agent that then asks a database agent for further information on the vehicle. While there is normally no human involved in these exchanges, using CE as a uniform information representation avoids communication problems — the meaning of human-provided information is not changed by some translation process to a different formal language — while also making it easier for humans to audit the exchanges when necessary. Also, on occasion, it will be useful to copy selected messages to a human for information.

To summarise, our main requirements are to support the following kinds of conversational interactions:

- NL to CE query or CE facts (a ‘confirm’ interaction);
- CE query to CE facts (an ‘ask-tell’ interaction);

- exchange of CE facts (a ‘tell’ interaction);
- gist CE to full CE (an ‘expand’ interaction);
- CE to CE rationale (a ‘why’ interaction).

In the following section, we formalise these kinds of conversational interactions by identifying a set of conversational primitives and valid interaction sequences.

III. Controlled English conversation cards (CE-Cards)

We conceptualise a conversation as a series of *cards* exchanged between agents, including humans and software services. Each card contains text, which can be natural (NL) or controlled (CE) language. To support human-machine conversation we allow three kinds of card content: fully-natural language, formal Controlled English, and a form of template-based CE that provides the gist of complex sets of CE sentences for brevity and easier human-readability. Drawing upon software agent research, a conversation unfolds through a series of primitive *communicative acts*; for example, queries, assertions, or requests [3, 4]. The key difference in our work is that we need communicative acts to support not only machine→machine communication, but also human→machine and machine→human.

III.A. CE-Cards model

Based on our requirements, we model several sub-types of card, shown in Figure 1 and given in CE form in the appendix. The three direct sub-types of card — CE card, NL card and gist card — provide important context for their content because it is not possible to unambiguously determine whether a piece of text is NL, CE, or gist by parsing it. For example, compare the NL sentence “there is a person named John” with the CE statement “there is a person named p1234 that is known as John”. If the parser interprets ‘John’ as an identifier then the first sentence could be misinterpreted as CE. (Note however that it is possible to determine that a string is *not* CE if it fails to parse as CE, in which case it could be NL or a gist.)

We define the following sub-types of CE card, each corresponding to a particular communicative act:

ask card that contains a CE query;

tell card that contains CE statements other than queries (e.g. facts or rationale);

confirm card that contains CE content derived from the content of a preceding NL card;

expand card that requests the formal CE form of the content of a preceding gist card;

why card that requests an explanation (CE rationale) for the content of a preceding ask or tell card.

¹While not our main focus, *human→human* interactions are also supported via exchange of NL or CE messages.

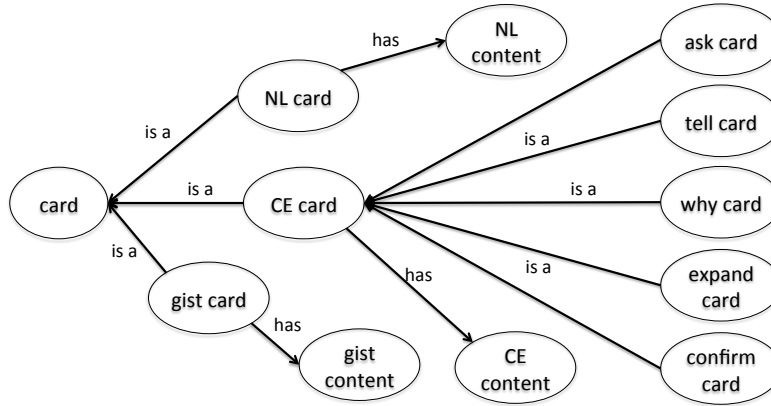


Figure 1: Graphical view of the CE-Cards model

CE ask and tell cards can be distinguished unambiguously by parsing their content (CE queries and facts respectively) so we don’t strictly need these sub-classes to be defined explicitly. However, having them makes it easier to specify conversational policies, as shown in Section III.B.

NL cards are not classified into “ask” and “tell” variants because their content is ambiguous. Without confirmation, it is not possible to determine for certain whether they represent a query or statement. Once the corresponding CE content is confirmed (via a CE confirm card) an ask or tell CE card is issued.

An expand card marks a transition from gist content to full CE; the content is able to specify CE entities that the sender wishes the expansion to focus on. For example, consider the following exchange²:

```

gist card: the red SUV is a threat
expand card: red SUV

tell card: there is a vehicle named v12345
           that has ‘red SUV’ as description and
           has XYZ456 as registration and...
  
```

Here, the agent issuing the expand card doesn’t want an expansion of “threat”, just the details of the SUV.

A why card marks a transition from CE facts to CE rationale; the content of a why card is able to specify CE entities that the sender wishes the explanation to focus on. For example:

```

tell card: there is a vehicle named v12345
           that is a threat and is located at
           central junction and...

why card: v12345 is a threat

tell card: v12345 is owned by HVT John
           Smith and...
  
```

Here, the sender of the why card wants an explanation of the threat as opposed to, say, the vehicle’s location.

An example instance of a card in CE syntax is shown below.

²For brevity, this and the next example are shown in pseudocode rather than complete CE format.

```

there is a tell card named ‘#2b’ that
is from the agent tasker and
is to the agent broker and
is in reply to the card ‘#2a’ and
has content the CE content
‘there is an HVT sighting named h00453 that
has the vehicle v01253 as target vehicle and
has the person p670467 as hvt candidate’.
  
```

This is a tell card from an agent called tasker to another agent called broker, reporting a high value target sighting. The card is a response to a previous card: all cards have unique identifiers, allowing conversation “threads” to be identified. The example shows the use of various card *attributes*, defined formally as CE relationships in the appendix. Every card is from some individual human or software agent. A card is to either an individual agent or a named group (e.g. a team in the MOSN context); a card can be to multiple recipients. In addition to the attributes shown here, every card has a timestamp (the UTC for when the card was sent, from the sender’s point-of-view) and may optionally have one or more linked resources, for example an associated image, video or audio stream, or document.

The three sub-concepts of card — CE card, gist card, NL card — can thus be defined in terms of restrictions on their content. For example, a CE card is a card that has content which is CE content. However, having explicit named sub-concepts makes the definition of conversational policies simpler, as shown below.

III.B. CE-Cards conversation policies

A *conversation* is a sequence of cards exchanged between two or more agents, with causal relationships between each pair of consecutive cards in the sequence (usually denoted by the identifier of the preceding card being used as the value of the succeeding card’s *is in reply to* attribute). Following [4], we define *conversational policies* as rules that describe permissible conversations between agents, specifying allowed sequences of cards and constraints on the attributes and content of individual cards depending on their place in a sequence. Figure 2 sketches the set of sequence rules for the card types defined in our model. A full discussion of the constraints on card attributes and concepts

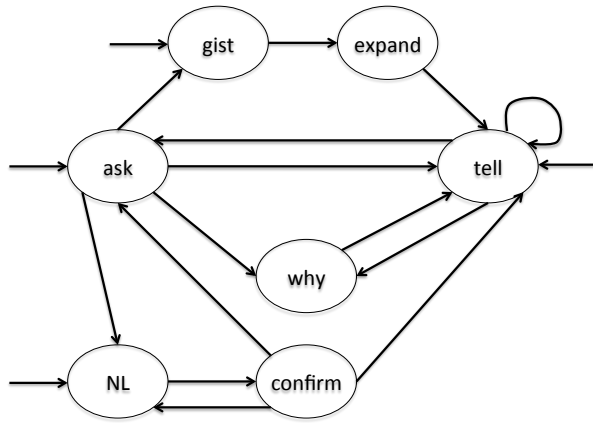


Figure 2: Conversation sequence rules for CE-Cards

accompanying this sequence is outside the scope of this paper, but examples are provided below and in Section IV.

In terms of our requirements for CE-Cards, the key interactions in the sequence in Figure 2 are as follows:

- The most basic form of conversation is an exchange consisting of an ask card a followed by a tell card t where t is in reply to a and the content of t is expected to be CE statements that satisfy the CE query in a .
- A conversation initiated by a human will typically begin with an NL card n to a software agent which will attempt to process the NL content of n into CE and respond with a confirm card c containing either a CE query or CE statements (depending on how the NL was interpreted), where c is in reply to n . There are now three permitted responses to c :
 - the originating human agent may accept (or edit) the CE content and, if it is a CE query, issue this content in an ask card a , where a is in reply to c ;
 - the originating human agent may accept (or edit) the CE content and, if it consists of CE statements, issue this content in a tell card t , where t is in reply to c ;
 - the originating human agent may not accept the content and issue a (modified) piece of NL content in a new NL card n' , where n' is in reply to c .
- An agent may respond to an ask card with a template form of CE contained in a gist card g , to which the recipient may respond with an expand card e requesting the full CE form of the gist information. Now the recipient of e is expected to respond with a tell card t the contents of which are expected to be the full CE form of the contents of g (e is in reply to g , t is in reply to e).
- An agent may respond to a tell card t with a why card requesting an explanation for the contents of t ; the recipient of w is expected to respond with a tell

card t' , the contents of which are expected to be CE rationale for the contents of t (w is in reply to t , t' is in reply to w).

Conversation sequences are expected to begin with one of the following: an ask card, tell card, gist card, or NL card. More complex conversations can be constructed from the sub-sequences described above, and other permissible sequences. For example, following receipt of a tell card t , the recipient may issue a follow-up query in an ask card a , where a is in reply to t .

IV. Vignette and analysis

We use a surveillance vignette from [10] to provide an illustrative walkthrough of the use of our conversational model in a mission context. We analyse the initial steps of the vignette in terms of human-machine, machine-human, and machine-machine interactions, involving four interacting agents:

- a human officer on patrol in the field (*patrol*);
- a human intelligence analyst with an interest in investigating suspicious activities in the area (*analyst*);
- a software agent that mediates between humans and other agents (*broker*);
- a software agent that handles access to database and sensor resources (*tasker*).³

The interactions in the initial steps of the vignette are as follows, illustrated in Figure 3:

Step 1: The patrol reports a suspicious black saloon car, vehicle registration ABC123, moving south on North Road. The report is issued as a NL text message to the broker, which generates and confirms the CE form of the report with the patrol.

Step 2: The broker sends the patrol's report to the tasker, and a database query reveals that this vehicle is associated with a high value target, John Smith. This HVT sighting is communicated back to the broker.

Step 3: Based on its knowledge of mission priorities previously provided by the analyst, the broker issues a request to the tasker to track the location of the vehicle. An unmanned aerial vehicle (UAV) is assigned to this task.

Step 4: The UAV locates and tracks the black saloon as it heads south on North Road. The UAV reports that the vehicle stops near Central Junction. The analyst is alerted of this via the broker, and requests imagery from the UAV.

We now provide details of these conversational interactions using the CE-Cards model. Most of the following sequence of interactions is illustrated in Figure 4. For brevity we do not present exchanged cards in full CE syntax but instead use an abbreviated format as follows:

³Other configurations of the software agents are possible, for example splitting the *tasker* into multiple agents with responsibility for different kinds of resources; the aim here is to show machine-machine communication while keeping the scenario simple.

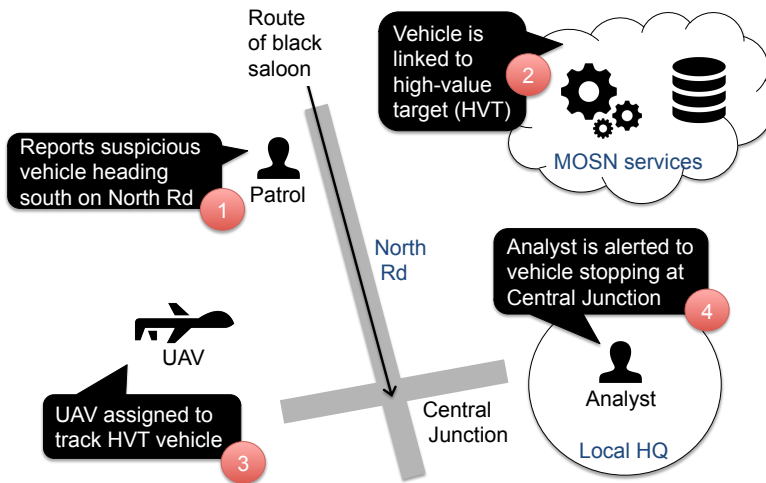


Figure 3: A surveillance vignette involving human and machine agents

<i>id.</i>	<i>card type</i>	<i>sender→recipient</i>	<i>in reply to id.</i>
<i>Content text</i>			
<i>Optional linked resource(s)</i>			

Step 1: Human patrol sends text message

#1a	NL	patrol→broker	
Suspicious vehicle driving south: black saloon car with license plate ABC123			

#1b	confirm	broker→patrol	in reply to #1a
there is a vehicle named v01253 that has 'black saloon car' as description and has black as colour and has saloon as body type and has ABC123 as registration.			

Additional information about location, direction and reporting patrol is also generated but not shown here.

#1c	tell	patrol→broker	in reply to #1b
CE as in card #1b: patrol confirms no change needed			

Step 2: Machine stores confirmed extracted facts

#2a	tell	broker→tasker	
CE as in card #1b			

#2b	tell	tasker→broker	in reply to #2a
there is an HVT sighting named h00453 that has the vehicle v01253 as target vehicle and has the person p670467 as hvt candidate.			

This statement is inferred CE that has been created as a result of fusing the new information from the patrol with background information already held in a database.

The recipient (or a human in a later forensic operation) could ask “why” to this response, in which case the rationale could be returned (not shown in Figure 4):

#2c	why	broker→tasker	in reply to #2b
CE as in card #2b			

#2d	tell	tasker→broker	in reply to #2c
because there is a person named p670467 that is known as 'John Smith' and is a high value target and the person p670467 has ABC123 as linked vehicle registration and there is a vehicle named v01253 that has ABC123 as registration.			

Step 3: Generation of sensing task to localize vehicle

A trigger is set in the system that will automatically create task instances whenever HVT sightings are reported.

#3a	tell	tasker→broker	
there is a task named t327893 that requires the intelligence capability localize and is looking for the vehicle v01253 and operates in the spatial area 'North Road' and is ranked with the task priority high.			

A CE description of the new task may be posted to the analyst for their information.

#3b	gist	broker→analyst	
A MALE UAV with EO camera has been tasked to localize a black saloon car (ABC123) with possible HVT John Smith in the North Road area.			

Assignments of sensing assets to tasks is done using the method described in [9], using a CE knowledge base of suitable sensor and platform types for a range of intelligence, surveillance, and reconnaissance tasks. The analyst could request an expansion of the above gist by means of an expand card; the expansion would be expressed in terms of the CE knowledge base, to justify that choice of asset (see Figure 4; messages not shown here for space reasons).

Step 4: Tracking updates are reported to the analyst

Here, there are a number of possibilities depending on how closely the analyst wishes to follow the tracking of the

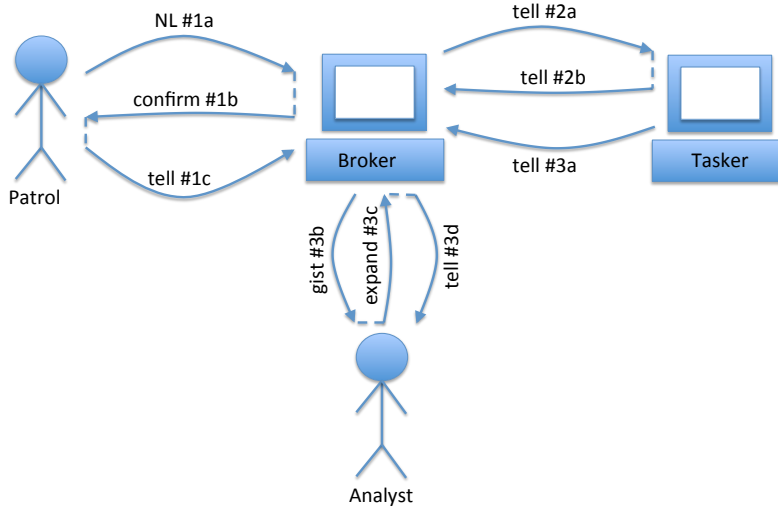


Figure 4: Interactions for steps 1–3 of the vignette

black saloon. This would be handled by the analyst expressing preferences to the broker via ask cards. For simplicity, we assume the analyst wishes to be alerted when the vehicle stops at a location:

#4a	gist	broker→analyst	
Vehicle ABC123 with possible HVT John Smith has stopped at location Central Junction.			
<i>Link to map showing position of vehicle</i>			

At this point the analyst may request imagery from the UAV:

#4c	NL	analyst→broker	
Show me live imagery from the UAV.			

There will now be a confirmation conversation to determine that this is a CE query, and an ask card issued, to which the broker will respond with a tell card including a link to the imagery as a resource attribute. Details of these interactions are similar to Step 1.

The above analysis illustrates most of the sub-sequences in Figure 2, and shows that the CE-Cards model is sufficient to support interactions among human and software agents in an MOSN context. The model has been designed to be minimal in terms of our requirements to support conversational flows from natural (NL) to controlled (CE) language, and back. The seven main types of card can be grouped in terms of which parts of the flow they support: NL→CE (NL, confirm), CE→CE (ask, tell, why), CE→NL (gist, expand).

V. Prototype implementation

We have implemented a prototype architecture based around a *CE-card server* which supports exchange of CE-cards among human or machine parties, while acting also as a repository for the cards exchanged. Figure 5 illustrates this server communicating with two agents: a user’s agent (e.g. the patrol in our vignette), and the broker. Soft-

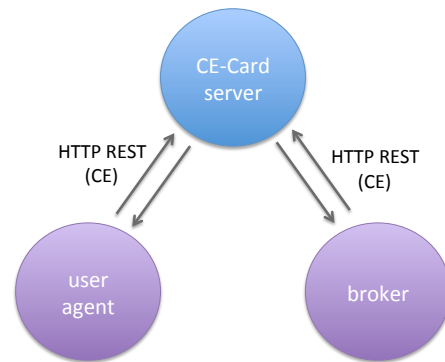


Figure 5: Illustration of the prototype architecture showing two example agents communicating via the CE-card server

ware agents representing human and machine parties can be implemented in any programming language. In the current prototype, the user agent is implemented as a web app (running on any browser) which is able to communicate with the CE-card server. Note that the CE-card server exchanges cards which are represented directly in CE, and that the CE-card server stores the same cards as pure CE statements. This means that CE is used uniformly as internal representation, transport layer, and user representation language. The architecture can be deployed on a service-oriented architecture for mission support such as the ITA Sensor Fabric [12].

Processing of CE information, including automated reasoning of the kind exemplified by the broker and tasker agents in Section IV, is performed by a component of the prototype called the Controlled English Store (or “CE Store”) — a research-grade implementation of a pure CE processing environment, developed to support ongoing research, experimentation and evaluation using the CE language. In addition to processing CE sentences it also provides a pluggable agent infrastructure within which machine agents can be easily added. The CE Store provides basic capabilities for model definition, simple workflow execution, definition of knowledge (facts), queries, rules

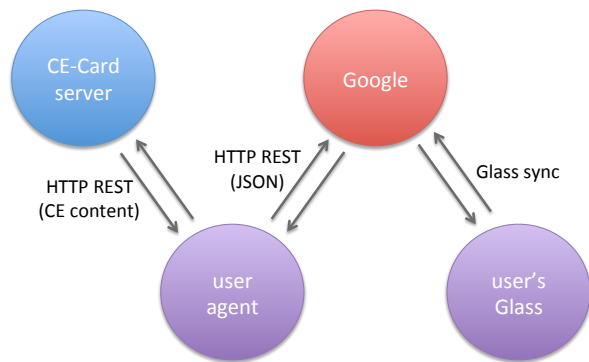


Figure 6: Illustration of the prototype architecture integrating the simulated Google Glass Mirror API

and annotations. The handling of the conversation policies within the prototype implementation of the Section IV vignette is performed by the CE Store also.⁴

As a concept-of-use, we created a prototype user interface based on the Google Glass Mirror API structure.⁵ At the time, Google’s API was not publicly accessible, but following its main structure allowed us to explore this type of interface for users in the field. In particular, we were interested in how a speech-based interface in conjunction with an eyeline-mounted display can be used to receive NL input and feed back short CE sentences and our gist format (we would envisage longer CE text being directed to a user’s handheld device). Integration of our simulated version of the Mirror API with the CE-card server architecture is illustrated in Figure 6. Figure 7 shows a screenshot of the Glass-style interface operating with a head-mounted camera and speech input to a laptop computer. This Glass-style interface was implemented by extending and making dynamic the Glassim simulator.⁶ The screenshot shows the user’s view, with the simulated “eyeline” display overlaid onto the live video from the head-mounted camera (while the user was observing a road).⁷

The prototype implementation, along with the Glass-style concept-of-use, has been evaluated informally by subject-matter experts from the US Army Research Laboratory and UK Ministry of Defence. The way in which the system “plays back” natural language as CE was highlighted as a particularly beneficial feature. Work is now underway to conduct more formal experiments with human subjects working in collaboration with software agents using NL, CE, and the template-based gist format as explained previously.

VI. Discussion

Research in agent communication languages (ACLs) [3, 4] viewed conversations as sequences of communicative acts,

⁴A publicly-available version of the CE Store (known formally as the IBM Controlled Natural Language Processing Environment) is available for download at: <http://ibm.co/RDIa53>

⁵<http://developers.google.com/glass/>

⁶<http://glassim.com/>

⁷Demonstration videos of the Section IV vignette can be viewed at: <http://www.usukitacs.com/node/2552>



Figure 7: Prototype implementation of Google Glass-style concept-of-use

drawing on work in philosophical linguistics. The idea of *illocutionary acts* from speech act theory [1] was adopted as a basis for ACL messages having explicit “performatives” that classify messages as, for example, assertives (factual statements), directives (such as requests or commands), or commissives (that commit the sender to some future action).

Our model features speech act-style performatives only for CE→CE interactions (ask and why are directives, tell is an assertive), as these support machine-machine communication. However, because CE is machine-processable, in principle the receiver could determine the illocutionary act from the message content. This is already true for ask and tell (CE queries versus CE facts); there is currently no CE form for a “why” query but one could be added to the language. In our approach, NL and gist cards do not have explicit performatives because the illocutionary act is determined by the human sender or receiver. The purpose of the confirm card is to disambiguate the intended act to allow software agents to respond as expected; the purpose of a gist card is to make complex CE easier for a human to understand and determine the sender agent’s intent (e.g. assertive or directive).

The tasker agent incorporates the results on previous work in resource allocation in MOSNs, where a knowledge-based system matches sensing assets to mission tasks [9]. Because this system is essentially performing the role of a “facilitator” in software agent research [4], a future possibility is to extend the CE-Cards model to support “brokerage” acts such as advertisements, subscriptions, or contracts.

The processing of NL cards to extract the information in a CE format builds upon ongoing research in information extraction using CE [13]. The main difference between that research and the usage in this context is the increased dependence on lexical descriptions for the concepts, rela-

tionships and attributes within the CE model to better enable the detectability of phrases and terms within NL statements and questions. The high-level approach taken is to first shallow-parse the NL text into component words and phrases and to seek these within the current set of available CE models available to the processing agent. If suitable matches are not detected using this simplistic approach then the NL sentence is sent off to the traditional NL processing using full lexical parsing of the sentence to determine whether this additional lexical knowledge can provide further accuracy. In all cases (including partial parses) the successfully extracted information from the NL sentence is converted to CE and returned to the user for review (if policy allows) and correction in any subsequent response.

The generation of gist messages is currently based mainly on the use of pre-defined templates for different parts of the CE model where simple variable substitution is used to populate the templates against the actual data for a given situation (an example of such a template is given in cards #3b and #4a in the walkthrough in Section IV). The templates can be used individually or combined as fragments to form a larger summary when the relevant information spans multiple templates. CE statements regarding the mapping of these templates and the relative importance of concepts, relationships and attributes are defined in the language of the CE meta-model. This builds on a technique known as *linguistic transformation* [6] whereby the information required to undertake linguistic transformations such as summaries is communicated directly in the CE language. Future research may look to integrate more advanced summarisation algorithms into this CE-based environment to make the summary generation capability more closely matched to human readability and relevance expectations.

One area of active experimentation is the use of more flexible forms of `confirm` card. Our original intent was to use the confirm card as a means to feed-back CE content to the user, to allow them to confirm the acceptability of the machine's interpretation of their original NL input. Under certain circumstances however, we see potentially-considerable value in supporting other forms of "confirm" feedback. For example:

- There may be cases where information management policy dictates that "verbatim" CE feedback should not be provided to the sender. This is especially true in a multi-agency situation, where there may be issues of confidentiality between collaborating parties [8]. In such cases, it would be useful for the receiver of the NL message to acknowledge receipt (e.g. "Thank you for your message") without revealing their precise interpretation of the information (which could expose, for example, inferred or extracted knowledge, or simply elements of their information model they would wish to keep confidential). Of course, this means that there is no formal confirmation by the sender of their intended meaning so, as discussed above, confirmation may need to be established by a (human) third party.

Such cases can easily be handled by a variant of the

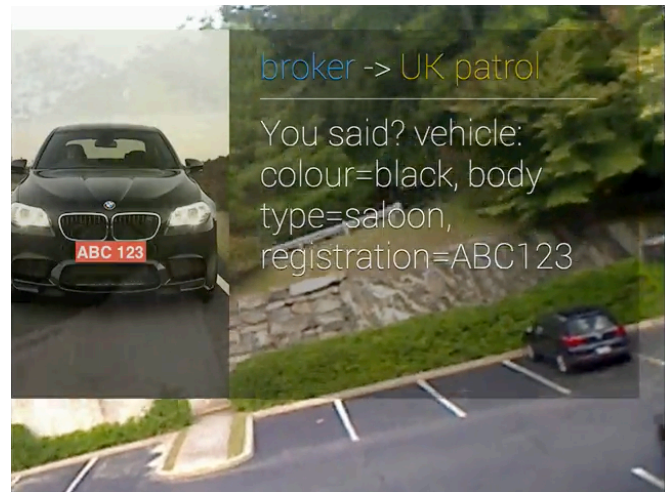


Figure 8: Experimentation with a graphical form for confirm or gist cards

`confirm` card that permits NL content (e.g. "Thank you for your message"), and does not oblige the recipient to respond.

- Another area for experimentation is the use of a graphical form for `confirm` card content, which our early trials suggest may be particularly useful in a form factor such as the Glass-style eyeline-mounted display. An example of this is shown in Figure 8 where, instead of a CE version of card #1b in Section IV, the user sees a combination of machine-generated image and compact text. The style/format (e.g. text and/or graphics) could be determined based on additional contextual factors such as the user, their role, the current operational tempo and the form factor of the device they are using.

We regard this kind of flexible structured content as an extension of our original notion of "gist" content; indeed, we envisage many cases where the system could use this kind of enriched graphical/text content in a `gist` card also. In such cases, however, if the user then requested an expansion of some specified elements of the "enriched gist", the content of the corresponding `expand` card would also need to be "enriched gist".

In conclusion, our initial experimentation with the prototype leads us to propose two changes to the original model in Figure 1, to support the above considerations:

- The potential utility of non-CE forms of confirmation message leads us to model the `confirm` card as a 'generic' kind of card, open to any kind of content: NL, CE, or gist.
- The value in allowing more flexible graphical/text content in both `gist` and `expand` cards, leads us to model the `expand` card as a kind of `gist` card.

The proposed revised model is shown in Figure 9. Colouring is used here to more clearly show the different

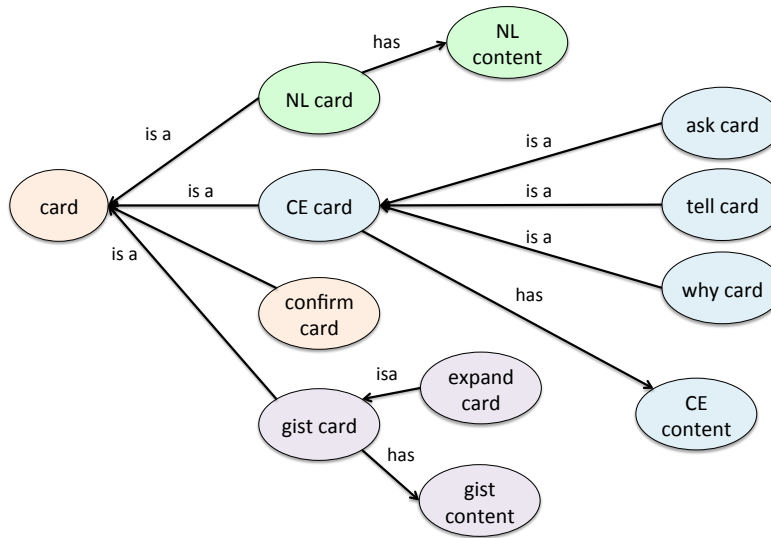


Figure 9: Proposed revised CE-Cards model, based on initial experimentation

kinds of card and content (NL, CE, gist, and ‘generic’). The model also allows for a ‘generic’ card to have multiple types of content so, for example, a single confirm card could carry both gist content and CE content, allowing the receiving agent greater flexibility in how it processes that content. For example, a user agent with a Glass-style interface may use gist content as in Figure 8, while a user agent on a tablet device may render the CE content (or offer the user both options).

VII. Conclusion and future work

This paper has introduced a model to support human-machine conversational interactions in a mission-oriented sensor network context, and shown how the model can be applied in practice. A key focus of our future work is developing these ideas in the multi-agency mission context, where information and assets are shared among different coalition partners with varying levels of trust, and conversations involve negotiations over access to resources. One area of ongoing work is in researching the effectiveness of Controlled English as a representation medium for security and resource management policies, that are both machine-processable (enforceable) and human-understandable. Initial work in this area is reported in [7] and future work will involve integrating that work into the conversational context.

A second major focus of the work going forward is to provide support for trade-offs relating to quality-of-information (QoI). QoI in multi-agency mission support networks is highly variable, both due to the nature of the sources and the network capacity. Examples of QoI issues in the context of our vignette include:

- *QoI issues stemming from the role of humans as sources of information:* For example, in our vignette we assume the officer on patrol is a trained observer and it is likely that the word “suspicious” (card #1a, Section IV) has a codified meaning and can be directly

mapped to some kind of certainty range. However, if that message came from an untrained or unknown individual, or via social media, then the subjective meaning of “suspicious” would need to be estimated and the certainty recorded.

- *Quality of information retrieved from database sources:* For example, information obtained by the tasker agent in the vignette (cards #2b and #2d, Section IV) may have associated uncertainty that the vehicle is associated with HVT John Smith (the information may be out of date or inaccurate).
- *Quality of information from physical sensors:* When the UAV attempts to track the black saloon, the accuracy of its observations will be subject to known or unknown QoI parameters. For example, observations such as the one that lead to card #4a in Section IV depend on the ability of the UAV system to accurately localise and track that particular vehicle.

These various forms of uncertainty may utilise different representations, for instance probabilities vs symbolic ratings, making information fusion a challenge. Going forward, we anticipate that the use of rationale, provenance, subjective logic, and gist (to convey summaries to a human user) can help. Note that we expect the fusion/understanding of all these sources of uncertainty to be fundamentally a human activity with support from the machine to make understanding/audit easier (through gist and rationale).

VIII. Acknowledgments

This research was sponsored by the US Army Research Laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted

as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the US Government, the UK Ministry of Defence or the UK Government. The US and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

We thank our collaborators in the Network and Information Sciences International Technology Alliance (ITA, <http://www.usukitacs.com>) especially Tom La Porta (Penn State University, USA), Simon Bray (Dstl, UK), Paul Sullivan (ARL, USA), and attendees at the 2013 ITA Annual Fall Meeting, Palisades NY, USA. We also thank the reviewers and participants in the MiSeNet'13 workshop at MobiCom 2013, Miami FL, USA.

References

- [1] J. Austin and J. Urmson. *How to Do Things With Words*. Harvard University Press, 1975.
- [2] B. Broome. Data-to-decisions: a transdisciplinary approach to decision support efforts at ARL. In *Proc Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR III (SPIE Vol 8389)*. SPIE, 2012.
- [3] Foundation For Intelligent Physical Agents. FIPA communicative act library specification, 2002.
- [4] Y. Labrou and T. Finin. Semantics and conversations for an agent communication language. In M. N. Huhns and M. P. Singh, editors, *Readings in agents*, pages 235–242. Morgan Kaufman, 1998.
- [5] D. Mott. Summary of ITA Controlled English, 2010.
- [6] D. Mott and J. Hendler. Layered controlled natural languages. In *3rd Annual Conference of the International Technology Alliance (ACITA)*, 2009.
- [7] C. Parizas, D. Pizzocaro, A. Preece, and P. Zerfos. Managing isr sharing policies at the network edge using controlled english. In *Proc Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IV (SPIE Vol 8742)*. SPIE, 2013.
- [8] T. Pham, G. Cirincione, D. Verma, and G. Pearson. Intelligence, surveillance, and reconnaissance fusion for coalition operations. In *Proc 11th International Conference on Information Fusion*, 2008.
- [9] A. Preece, D. Pizzocaro, D. Braines, and D. Mott. Tasking and sharing sensing assets using controlled natural language. In *Proc Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR III (SPIE Vol 8389)*. SPIE, 2012.
- [10] A. Preece, D. Pizzocaro, D. Braines, D. Mott, G. de Mel, and T. Pham. Integrating hard and soft information sources for D2D using controlled natural language. In *Proc 15th International Conference on Information Fusion*, 2012.
- [11] J. Sowa. *Common Logic Controlled English*, 2004.
- [12] J. Wright, C. Gibson, F. Bergamaschi, K. Marcus, R. P. K. Marcus, R. Pressley, G. Verma, and G. Whipps. A dynamic infrastructure for interconnecting disparate ISR/ISTAR assets (the ITA Sensor Fabric). In *IEEE/ISIF Fusion 2009 Conference*, June 2009.
- [13] P. Xue, D. Mott, D. Braines, S. Poteet, A. Kao, C. Giammanco, T. Pham, and R. McGowan. Information extraction using controlled english to support knowledge-sharing and decision-making. In *17th IC-CRTS “Operationalizing C2 Agility”*, 2012.

Appendix: A Brief Guide to CE

CE is used to define both models and instances. Model definitions take the form of concept definitions. CE conceptualise sentences are intended to define concepts by example; that is, they provide generalised examples of how to say things about concepts, including relationships between them. A CE model may also include the definition of logical inference rules which are used to express further information about the concepts and relationships and how they are logically related. Concepts may be specialisations of other concepts, indicated by *is a* declarations. Relationships may be defined between concepts (for example, the relationship *is to* between the concepts *card* and *agent*). The following definitions cover the core CE-Cards model (as presented in Figure 1; the proposed revised version is not shown here):

```
conceptualise a ~ card ~ C that
  has the timestamp T as ~ timestamp ~ and
  has the resource R as ~ resource ~.
conceptualise the card C
  ~ is from ~ the individual I and
  ~ is to ~ the agent A and
  ~ is in reply to ~ the card Q.
conceptualise a ~ CE card ~ C that
  is a card and
  has the CE content C0 as ~ content ~.
conceptualise a ~ gist card ~ C that
  is a card and
  has the gist content C0 as ~ content ~.
conceptualise an ~ NL card ~ C that
  is a card and
  has the NL content C0 as ~ content ~.
conceptualise an ~ ask card ~ C
  that is a CE card.
conceptualise a ~ confirm card ~ C
  that is a CE card.
conceptualise a ~ expand card ~ C
  that is a CE card.
conceptualise a ~ tell card ~ C
  that is a CE card.
conceptualise a ~ why card ~ C
  that is a CE card.
```