Conversational Homes: A Uniform Natural Language Approach for Collaboration Among Humans and Devices

Dave Braines, Nick O'Leary, Anna Thomas
Emerging Technology,
IBM United Kingdom Ltd,
Hursley Park, Winchester, UK
Email: {dave_braines, nick_oleary, annaet}@uk.ibm.com

Daniel Harborne, Alun Preece, Will Webberley
Crime and Security Research Institute /
School of Computer Science and Informatics,
Cardiff University, Cardiff, UK
Email: {HarborneD, PreeceAD, WebberleyWM}@cardiff.ac.uk

Abstract—**As devices proliferate, the ability for us to interact with them in an intuitive and meaningful way becomes increasingly challenging. In this paper we take the typical home as an experimental environment to investigate the challenges and potential solutions arising from ever-increasing device proliferation and complexity. We describe and evaluate a potential solution based on conversational interactions between "things" in the environment where those things can be either machine devices or human users. Our key innovation is the use of a Controlled Natural Language (CNL) technology as the underpinning information representation language for both machine and human agents, enabling humans and machines to trivially "read" the information being exchanged. The core CNL is augmented with a conversational protocol enabling different speech acts to be exchanged within the system. This conversational layer enables key contextual information to be conveyed, as well as providing a mechanism for translation from the core CNL to other forms, such as device specific API (Application Programming Interface) requests, or more easily consumable human representations. Our goal is to show that a single, uniform language can support machine-machine, machine-human, human-machine and human-human interaction in a dynamic environment that is able to rapidly evolve to accommodate new devices and capabilities as they are encountered. We also report results from our first formal evaluation of a Conversational Homes prototype and demonstrate users with no previous experience of this environment are able to rapidly and effectively interact with simulated devices in a number of simple scenarios.**

Keywords–IoT; Controlled Natural Language; Conversational Interaction.

I. Introduction

From an individual agent's perspective, the Internet of Things (IoT) can be seen as an increasingly large and diverse world of other agents to communicate with. Humans are agents too in this world, so we can observe four kinds of communication: (i) human-machine, (ii) machine-human, (iii) machine-machine, and (iv) human-human. There is a tendency to consider human-oriented (i, iv) and machine-oriented (ii, iii) interactions as naturally requiring different kinds of communication language; humans prefer natural languages, while machines operate most readily on formal languages. In this paper, however, we consider what the IoT world might look like where humans and machines largely use a common, uniform language to communicate. Our design goal is to support communication activities such as: the discovery of other agents and their capabilities, querying other agents and receiving understandable information from them, and obtaining rationale for an agent's actions. The proposed approach must be able to cope with rapid evolution of an IoT environment that needs to accommodate new devices, capabilities, and agent types. In Section II, we consider why human users might find such an environment more appealing when machines communicate using an accessible and human-friendly language, than when machines use a traditional machine-to-machine formalism. Section III substantiates our proposed approach using a series of vignettes, while Section IV presents evidence that human-machine and machine-machine interactions can be facilitated via a CNL communication mechanism as well as a full description and analysis of the recent initial Conversational Homes evaluation study. Section V concludes the paper.

This paper extends ideas first proposed in [1], specifically by reporting on the first formal evaluation of the conversational protocol described in that work. Sections I–III in this paper are largely unchanged from [1], with Section IV being substantially expanded to describe the evaluation setup and results. Section V is also updated to reflect these latest developments and our plans for future work.

II. Background and Related Work

A key part of our approach is to consider how humans "want" to interact with machines in the world. To help us gain insights into these latent human requirements we look towards existing trends and events occurring in the world and use these as inspiration to help us form our hypotheses about what a conversational environment for human-machine agents might entail. For example, in this work we consider recent interest in conversational technologies such as chatbots [2], conversational computing [3], and conversational agents [4]. The remainder of this section covers this human-motivated perspective and develops ideas first presented in [5].

A. Social Things

The advent of Twitter as a means of social communication has enabled a large number of otherwise inanimate objects to have an easily-accessible online presence. For example, Andy Stanford-Clark created an account for the Red Funnel ferries that service the Isle of Wight in the UK. The account [6] relays real-time information about the ferry arrivals and departures allowing a subscriber of the account to see if they are running on time.

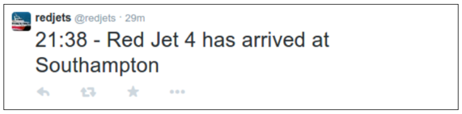

Figure 1: Redjet tweet example.

Another similar example is an unofficial account for London's Tower Bridge [7]. Its creator, Tom Armitage, created a system that took the published scheduled of bridge opening and closing times and produced a Twitter account that relayed that information.

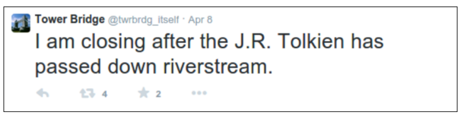

Figure 2: Tower Bridge tweet example.

A key difference between the ferries and the bridge accounts is that the ferries are just relaying information, a timestamp and a position, whereas the bridge is speaking to us in the first-person. This small difference immediately begins to bring a more human nature to the account. But, they are ultimately simple accounts that relay their state to whomever is following them, providing an easily consumable feed of information on an existing platform.

This sort of thing seems to have caught on particularly with the various space agencies. We no longer appear able to send a robot to Mars, or land a probe on a comet without an accompanying Twitter account bringing character to the events. The Mars Curiosity Rover has had an account [8] since July 2008 and regularly shares images it has captured. There's always a sense of excitement when these inanimate objects start to have a conversation with one another. The conversations between the European Space Agency Philae lander [9] and its Rosetta orbiter [10], as the former began to lose power and had to shutdown, generated a large emotional response on social media. The lander, which was launched into space years before social media existed, chose to use its last few milliamps of power to send a final goodbye.

The reality, of course, is that the devices did not create these tweets. Communication with them remains the preserve of highly specialized engineers, and their personalities are a creation of their public relations agencies on this planet. There are however, examples of machine participation on social media provided by social bots [11]. On occasion, these entities can masquerade as human agents and alter the dynamics of social sense-making and social influence.

B. *Seamlessness vs Seamfulness*

The IoT makes possible a future where our homes and workplaces are full of connected devices, sharing their data, making decisions, collaborating to make our lives better [12]. Whilst there are people who celebrate this invisible ubiquity and utility of computing, the reality is going to be much more complicated.

Mark Weiser, Chief Scientist at Xerox PARC, coined the term "ubiquitous computing" in 1988 as recognition of the changing nature of our interaction with computers [13]. Rather than the overt interaction of a user sitting in front of a computer, ubiquitous computing envisages technology receding into the background of our lives.

Discussion of ubiquitous computing often celebrates the idea of seamless experiences between the various devices occupying our lives. Mark Weiser advocated for the opposite; that seamlessness was undesirable and a self-defeating attribute of such a system. He preferred a vision of "Seamfulness, with beautiful seams" [14].

The desire to present a single view of a system, with no joins, is an unrealistic aspiration in the face of the cold realities of Wi-Fi connectivity, battery life, system reliability and the status of cloud services. Presenting a user with a completely monolithic system gives them no opportunity to connect with, and begin to understand, the constituent parts. That is not to say all users need this information all of the time, but there is clearly utility to some users some of the time: when you come home from work and the house is cold, what went wrong? Did the thermostat in the living room break and decide it was the right temperature already? Did the message from the working thermostat fail to get to the boiler? Is the boiler broken? Did you forget to cancel the entry in your calendar saying you'd be late home that day? Without some appreciation of the moving parts in a system, a user cannot feel any ownership or empowerment when something goes wrong with it. Or worse yet, how can they avoid feeling anything other than intimidated by this monolithic system that responds in a manner akin to, "I'm Sorry Dave, I'm afraid I can't do that".

This is the justification for beautiful seams: they help you understand the edges of a device's sphere of interaction, but should not be so big as to trip you up. For example, such issues exist with the various IP connected light bulbs that are available today. When a user needs to remember what application to launch on their phone depending on what room they are walking into and what manufacturer's bulbs happen to be in there, the seams have gotten too big and too visible.

Designer Tom Coates has written on these topics [15]. He suggests the idea of having a chat-room for the home:

"Much like a conference might have a chat-room so might a home. And it might be a space that you could duck into as you pleased to see what was going on. By turning the responses into human language you could make the actions of the objects less inscrutable and difficult to understand..."

This relates back to the world of Twitter accounts for Things, but with a key evolution. Rather than one-sided conversations presenting raw data in a more consumable form, or Wizard-of-Oz style man-behind-the-curtain accounts, a chat-room is a space where the conversation can flow both ways; both between the owner and their devices, and also between the devices themselves.

C. *Getting Things Communicating*

For devices to be able to communicate they need to share a common language. Simply being able to send a piece of data across the network is not sufficient. As with spoken language, the context of an interaction is important too.

This model of interaction applies to both the data a device produces, as well as the commands it can consume. There are a number of technologies available for producing such a shared model. For example: HyperCat [16], a consortium of companies funded by the UK Government's innovation agency in 2014. It provides a central catalog of resources that are described using RDF-like triple statements. Each resource is identified by a URI allowing for ease of reference. URIs are a key component in building the World Wide Web and are well understood, but they are a technology used primarily by computers. They do not provide a human-accessible view of the model.

Furthermore, to enable a dynamic conversation, any such model needs to be adaptable to the devices that are participating, especially when one of those participants is a human being.

D. Talking to Computers

The most natural form of communication for most humans is that of their own spoken language, not some JSON or XML encoded format that was created with the devices as the primary recipient. Technical specialists can be trained to understand and use technical machine languages, but this overhead is not acceptable to more casual everyday users who may wish to interact with the devices in their home. In addition to this, we are living in an age where talking to computers is becoming less the preserve of science fiction: Apple's Siri, OK Google, Microsoft Cortana all exist as ways to interact with the devices in your pocket. Amazon Echo exists as a device for the home that allows basic interaction through voice commands. This means that there is now a plausible expectation that an everyday person could interact with complex devices in their home in a natural conversational manner.

Natural Language Processing (NLP) is one of the key challenges in Computer Science [17]. In terms of speech understanding, correctly identifying the words being spoken is relatively a well-solved problem, but understanding what those words mean, what intent they try to convey, is still a hard thing to do.

To answer the question "Which bat is your favorite?" without any context is hard to do. Are we talking to a sportsperson with their proud collection of cricket bats? Is it the zookeeper with their colony of winged mammals? Or perhaps a comic book fan is being asked to choose between incarnations of their favorite super hero.

Context is also vital when you want to hold a conversation. Natural language (NL) is riddled with ambiguity. Our brains are constantly filling in gaps, making theories and assumptions over what the other person is saying. For humans and machines to communicate effectively in any such conversational home setting, it is important that contextual information can be communicated in a simple, but effective, manner. This must be achieved in a manner that is accessible to both the human and machine agents in this environment.

E. Broader considerations

The focus of our research and the evaluation described later in this paper are exploring whether the use of a CNL technology can ease and/or speed the development of conversational systems, such as for an IoT enabled home. With this in mind we have not specifically attempted to build a system which has a rich or complex grammar or dialogue system, nor have we tried to create extensive, rich models or ontologies of the domain. There is no dialogue management required in our solution for this evaluation, and the required ontologies can be incredibly simple for the basic initial evalution activities.

Instead we have shown an approach where simple models can be quickly created by less technical users, to become the basis for systems such as the one evaluated in this paper. Our contribution to the literature is in showing an approach where the complexity and development time of the underlying models can be substantially reduced, ideally to a rate where real-time extensions can be made as new devices and capabilities are released. Such capabilities will be essential in any multi-organisation complex environment such as the IoT devices that could be used in a home environment.

We do acknowledge a rich body of research in the domain of multi-model interfaces [18] [19], dialogue systems and dialogue management [20] [21] [22] which would relate to the subsequent development of a complete system (regardless of whether it was underpinned by a CNL technology), but have not attempted to position our work against these for this particular evaluation.

There is also extensive literature on ontology engineering and the use of such systems in the context of dialogue [23] [24] [25] but again these are acknowledged but not specifically relevant to this simple evaluation against rapidly development CNL ontologies in a languaged aimed at less technical users.

III. Controlled Natural Language

To avoid a lot of the hard challenges of NLP, a CNL can be used. A CNL is a subset of a NL that uses a restricted set of grammar rules and a restricted vocabulary [26]. It is constructed to be readable by a native speaker and represents information in a structured and unambiguous form. This also enables it to be read and properly interpreted by a machine agent via a trivial parsing mechanism without any need for complex processing or resolution of ambiguity. CNLs range in strength from weaker examples such as simple style guides, to the strongest forms that are full formal languages with well-defined semantics. In our work, to identify a unifying language for both human and machine communication, we are focused on languages at the strong end of the scale, but we additionally wish to retain the requirement for maximal human consumability.

Ambiguity is a key issue for machine agents: whilst human readers can tolerate a degree of uncertainty and are often able to resolve ambiguity for themselves, it can be very difficult for a computer to do the same. CNLs typically specify that words be unambiguous and often specify the meaning that is allowed for all or a subset of the vocabulary. Another source of ambiguity is the phrase or sentence structure. A simple example is concerned with noun clusters. In English, one noun is commonly used to modify another noun. A noun phrase with several nouns is usually ambiguous as to how the nouns should be grouped. To avoid potential ambiguity, many CNLs do not allow the use of more than three nouns in a noun phrase.

There are two different philosophies in designing a CNL. As mentioned previously a weaker CNL can be treated as a simplified form of NL with a stronger CNL as an English version of a formal language [27]. In the case of a simplified form of NL, it can allow certain degrees of ambiguity in order to increase human accessibility. It relies on standard NLP techniques, lexical-semantic resources and a domain model to optimize its interpretation.

The alternative is to treat a CNL as an entirely deterministic language, where each word has a single meaning and no ambiguity can exist. Whilst computationally very efficient, it can be hard for a human user unfamiliar with the particular lexicon and grammar to write it effectively. This is because it competes with the user's own intuition of the language. The closer a CNL is to corresponding NL, the more natural and easy it is to use by humans, but it becomes less predictable and its computational complexity increases. The converse is also true. The more deterministic the CNL is, the more predictable it is, but the more difficult it is for humans to use.

In summary, in the operational setting described in this paper a CNL is designed to support both human usage and machine processing. It provides:

1) A user-friendly language in a form of English, instead of, for example, a standard formal query language (such as SPARQL or SQL). Enabling the user to construct queries to information systems in an intuitive way.
2) A precise language that enables clear, unambiguous representation of extracted information to serve as a semantic representation of the free text data that is amenable to creating rule-based inferences.
3) A common form of expression used to build, extend and refine domain models by adding or modifying entities, relations, or event types, and specifying mapping relations between data models and terminology or language variants.
4) An intuitive means of configuring system processing, such as specifying entity types, rules, and lexical patterns.

A good balance between the naturalness and predictability of the CNL is fundamentally important, especially to the human users as the strength and formality of the language increases.

A. An Introduction to ITA Controlled English

In previous research, we have proposed a specific CNL that is a variant of "Controlled English" known as ITA Controlled English, or just "CE" in shorthand [28]. This has been researched and developed under the International Technology Alliance (ITA) in Network and Information Science [29]. CE is consistent with First Order Predicate Logic and provides an unambiguous representation of information for machine processing. It aspires to provide a human-friendly representation format that is directly targeted at non-technical domain-specialist users (such as military planners, intelligence analysts or business managers) to enable a richer set of reasoning capabilities [30], [31].

We assert that CE can be used as a standard language for representation of many aspects of the information representation and reasoning space [32]. In addition to more traditional areas such as knowledge or domain model representation and corresponding information, CE also encompasses the representation of logical inference rules, rationale (reasoning steps), assumptions, statements of truth (and certainty) and has been used in other areas such as provenance [33] and argumentation [34].

In the remainder of this section we give a number of examples of the CE language. These are shown as embedded sentences in `this style`. All of these sentences are valid CE and therefore directly machine processable as well as being human readable.

The domain model used within CE is created through the definition of concepts, relationships and properties. These definitions are themselves written as CE conceptualise statements:

```
conceptualise a ~ device ~ D.
conceptualise an
   ~ environment variable ~ E.
```

These statements establish the concepts within the CE domain model enabling subsequent instances to be created using the same CE language:

```
there is an environment variable named
'temperature'.
```

A slightly more advanced example would be:

```
conceptualise a
   ~ controlling thing ~ C that
   is a device and
   ~ can control ~
     the environment variable E.
```

This defines "controlling thing" as a sub-concept of "device" and that it can have a "can control" relationship with an "environment variable". This therefore allows statements such as:

```
there is a controlling thing named
'thermostat' that
   can control the environment variable
   'temperature'.
```

In the latter conceptualise statement, "can control" is an example of a CE verb singular relationship. Functional noun relationships can also be asserted:

```
conceptualise a ~ device ~ D that
   has the value E as ~ enabled ~.
```

These two types of relationship construct allow a concept and its properties to be richly defined in CE whilst maintaining a strict subset of grammar. The use of verb singular and functional noun forms of properties provides a simple, but effective, mechanism to enable the conceptual model designer to use a language that is more natural and appealing to the human agents in the system.

The "is a" relationship used within conceptualise sentences defines inheritance of concepts, with multiple inheritance from any number of parents being a key requirement. It also allows any instance to be asserted as any number of concurrent concepts; an essential tool when attempting to capture and convey different contexts for the same information.

Whilst the examples given above are deliberately simplistic the same simple language constructs can be used to develop rich models and associated knowledge bases. The CE language has been successfully used in a wide range of example applications [35]. CE has been shown working with a reasonable number of concepts, relationships, queries and rules and has been used to model and interact with complex real-world environments with a high level of coverage and practical expressivity being achieved.

In our previous research into the application of the CE language we have observed that by gradually building up an operational model of a given environment, it is possible to iteratively define rich and complex semantic models in an "almost-NL" form that appeals to non-specialist domain users. For example, if the concept "device" was extended to include location information, the following query could be used to identify all devices of a particular type within a particular location:

```
for which D is it true that
   (the device D
     is located in the room V) and
   (the device D can measure
     the environment variable
        'temperature') and
   (the value V = 'kitchen').
```

Note that we do not expect casual users to write CE queries of this complexity; the later conversational interaction section will show how users can do this in a more natural form.

The model can be extended with rules that can be used to automatically infer new facts within the domain. Whenever such facts are inferred the CE language is able to capture rationale for why a particular fact is held to be true:

```
the room 'kitchen'
   is able to measure
     the environment variable
        'temperature' and
   is able to control
     the environment variable
     'temperature'
because
   the thermometer 't1'
     is located in the room 'kitchen' and
     can measure
        the environment variable
          'temperature' and
   the radiator valve 'v1'
     is located in the room 'kitchen' and
     can control
        the environment variable
          'temperature'.
```

From these basic examples you can see how the CE language can be used to model the basic concepts and properties within a given domain (such as an operating environment for IoT devices). Through assertion of corresponding instance data and the use of queries and rules it is possible to define the specific details of any given environment. It should also be clear to the reader that whilst human-readable the core CE language is quite technical and does not yet meet the aspiration of a language that would appeal to everyday casual users. The language itself can be improved, and as reported in earlier research there is the ability to build incrementally usable layers of language on top of the CE core language [36]. However, in addition to all of these potential advances in the core language there is also a key innovation that has been recently developed, which is to build a rich conversational protocol on top of the CE language [37]. This provides a mechanism whereby casual users can engage in conversation with a CE knowledge base using their own NL in a manner similar to human-human conversation.

B. Conversational Interaction

To enable a conversational form of CE, earlier research [38] has identified a requirement for a number of core interaction types based on speech-act theory:

1) A confirm interaction allows a NL message, typically from a human user, to be provided, which is then refined through interaction to an acceptable CE representation. This is useful for a human user who is perhaps not fully trained on the CE grammar. Through multiple such interactions, their experience builds and such interactions become shorter.
2) An ask/tell interaction allows a query to be made of the domain model and a well-formulated CE response given.
3) A gist/expand interaction enables the CE agent to provide a summary form ("gist") of a piece of CE, possibly adapted to a more digestible NL form. Such a gist can be expanded to give the underlying CE representation.
4) A why interaction allows an agent in receipt of CE statements to obtain rationale for the information provided.

This "conversational layer" is built within the core CE environment and is defined using the CE language. Within the CE model, these interactions are modeled as sub-types of the card concept.

```
conceptualise a ~ card ~ C that
   is an entity and
   has the timestamp T as ~ timestamp ~ and
   has the value V as ~ content ~ and
   ~ is to ~ the agent A and
   ~ is from ~ the agent B and
   ~ is in reply to ~ the card C.
```

The concept of an agent is introduced to represent the different parties in a conversation. This model provides a framework for such agents to interact by CE statements. By developing a conversational protocol using the CE language it enables the same language to be used for the domain in question (e.g., IoT devices in the home), as well as the act of communication. This means that agents with different operational domains can still communicate using a standard conversational model, so even if they cannot decode the items being discussed they are at least able to participate in the conversation. This idea is central to the proposed approach for conversationally enabled human and machine agents in an IoT context described in this paper.

C. Agent and ce-store interaction

In our ongoing experiments using the CE language we are able to define models, build knowledge bases, build machine agents and enable conversational interaction between them using some key components, which we will refer to here as ce-store. The Java-based implementation of the full ce-store [39] is publically available from github and an additional javascript-based version [40] is also available, specifically engineered to enable operation at the edge of the network, i.e., in a mobile browser environment.

For example, the domain model shown earlier in this paper is created through CE, (including the concepts, relationships and instances) and held within an instance of the ce-store, also referred to as a CE knowledge base. This store can either be maintained at a central point in the architecture, or distributed across systems through a federated set of separate ce-store instances. A centralized store provides a more straightforward system to maintain and ensures a single, shared model. Distributing the store allows for more localized processing to be done by the agents without having to interact with the system as a whole. Distributing the store also enables different agents to have different models, and for models to be rapidly extended "in the field" for only those agents that require those changes.

The choice of agent architecture influences how the store should be structured. When considering the types of conversation a chat-room for the home may need to support, there are two possible approaches.

1) The human user interacts with a single agent in the role of a concierge for the home. This concierge agent uses the CE knowledge base to maintain a complete situational awareness of the devices in the home and is able to communicate with them directly (see Figure 3). Interactions between concierge and devices do not use CE; only the concierge has a CE knowledge base.
2) The human user interacts with each device, or set of devices, individually. There may still be an agent in a concierge style role, but conversations can be directed at individual devices of interest as required (see Figure 4). Here, the concierge and all devices can communicate using CE and all have their own CE knowledge bases.

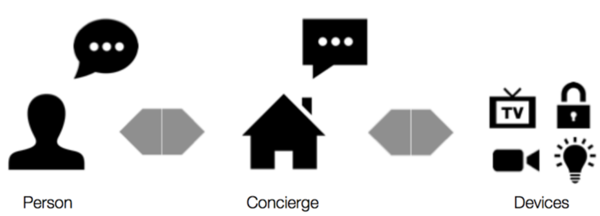

Figure 3: The human user interacts (via CE) only with the concierge.

Whilst the former would be sufficient to enable purely human-machine interaction, one of the goals of this work is to enable the human to passively observe the interaction of the devices in the home in order to help the human gain awareness of how the system is behaving. This will better enable the human user to see normal behavior over time and therefore prepare them for understanding anomalous situations when they arise.

Figure 4: The human user can interact (via CE) directly with all devices and with devices via the concierge.

As such, the latter approach is more suited for these purposes, perhaps with a concierge agent who is additionally maintaining the overall situation awareness from a machine-processing perspective.

D. Modelling the Conversation

In our proposed Conversational Homes setting there are a number of styles of interaction a human may wish to have with the devices in their home. This section considers four such styles and how they can be handled within a CE environment.

1) Direct question/answer exchanges: This is where a user makes a direct query as to the current state of the environment or one of the devices therein. For example: "What is the temperature in the kitchen?"

Through the existing conversational protocol and embedded simple contextual NL processing a machine agent is able to break down such a statement to recognize its intent. By parsing each word in turn and finding matching terms within the ce-store it can establish:

- it is a question regarding a current state ("What is ...")
- it is regarding the temperature environment variable instance
- it is regarding the kitchen room instance

At this point, the machine agent has sufficient information to query the ce-store to identify what devices in the model are in the right location and capable of measuring the required variable. If such a device exists, it can be queried for the value and reported back to the user. Otherwise, a suitable message can be returned to indicate the question cannot be answered, ideally conveying some indication of why not.

If the question is ambiguous, for example by omitting a location, the agent can prompt the user for the missing information. The concept of ambiguity for this kind of question is also captured in CE, for example by stating that for such an environment variable a location must be specified, perhaps even with a default location that can be assumed. With this knowledge available in CE the agent is able to determine that extra information is still required and can request this from the user as part of the conversation. The agent maintains information regarding the state of the conversation such that prompts can be made without requiring the user to repeat their entire question with the additional information included. By using the conversational protocol on top of the core CE language the human user and the device are able to converse in NL, for example:

User: *What is the temperature?*
Agent: *Where? I can tell you about the kitchen, the hall and the master bedroom.*
User: *The kitchen.*
Agent: *The temperature in the kitchen is 22C*

Other simple question types can be handled in this way, such as "where is…".

2) Questions that require a rationale as response: This is where a user requires an explanation for a current state of the system "Why is the kitchen cold?"

As with a direct question, an agent can parse the question to identify:

- it is a question asking for a rationale ("Why is …")
- it has a subject of kitchen
- it has a state of cold that, through the CE model, is understood to be an expression of the temperature environment variable.

To be able to provide a response, the model supports the ability to identify what can affect the given environment variable. With that information it can examine the current state of the system to see what can account for the described state. For example, "the window is open" or "the thermostat is set to 16C".

3) An explicit request to change a particular state: This is where a user, or a machine agent, makes an explicit request for a device to take an action "Turn up the thermostat in the kitchen"

To identify this type of statement, the model maintains a set of actions that can be taken and to what devices they can be applied. By incrementally matching the words of the statement against the list of known actions, a match, if it exists, can be identified. Further parsing of the statement can identify a target for the action.

```
conceptualise an ~ action ~ A that
  ~ is reversed by ~ the action B and
  ~ can affect ~ the controlling thing M.

if (the action A
  is reversed by the action B)
then (the action B
  is reversed by the action A).
```

The CE above demonstrates the ability to define a rule. These are logic constructs with premises and conclusions that get evaluated by the ce-store against each new fact added. Where a match in the premises is found, new facts are generated using the conclusions (with corresponding rationale). In this simple case it allows two-way relationships to be established without having to explicitly define the relationship in both directions.

```
there is an action named 'turn on'.
there is an action named 'turn off'.
the action 'turn on'
  is reversed by the action 'turn off'.
```

When a device receives an action, the trigger concept can be used to chain further sequences of actions that should occur. For example, when applied to a thermostat, the action "turn up" should trigger the action "turn on" to be applied to the boiler.

```
there is a trigger named ' tr1' that
  has 'turn up' as action and
  has 'boiler' as target device and
  has 'turn on' as target action.

the thermostat 'ts1'
  will respond to the trigger 'tr1'.
```

There is a natural point of contact here, with the popular 'If This Then That' framework (IFTTT) [41], specifically in that the use of conversational interactions could provide a nice way to implement IFTTT functionality. In future work we may consider the extent CE could be applied in IFTTT scenarios, and used to support a user-friendly form of programming for real-world objects, devices and situations.

4) An implicit desire to change a state: The styles considered so far have been explicit in their intent. There is another form whereby a statement is made that states a fact, but also implies a desire for an action to be taken.

This relies on Grice's Maxim of Relevance [42]. In the context of a conversation with the devices in a house, a statement such as "I am cold" should be taken as a desire for it to be warmer. The underlying information that can allow this Gricean inference to be implemented by machine agents using a simple algorithm is shown below:

```
there is a physical state
    named 'cold' that
  is an expression of
    the environment variable
      'temperature' and
  has 'warmer' as desired state.

there is a desired state
    named 'warmer' that
  has 'temperature' as target and
  has 'increase' as effect.
```

Once the intention of the statement has been identified, the store can be queried to find any actions that satisfy the requirement. These actions can then be offered as possible responses to the statement, or possibly automatically enacted.

Through these four simple dialogue examples we have demonstrated that through the use of a CE knowledge base and a set of machine agents using the conversational protocol a human user could carry out basic interactions with the devices in their home (human-machine). We have also shown how those devices convey key information back to the user, or ask follow on questions to elicit additional information (machine-human). These same interactions using the same CE language can be used to enable direct communications between machine agents regardless of human involvement (machine-machine). Whilst we have not explicitly demonstrated human-human communication it is clear that this can easily be supported within a system such as this, for example, by enabling different human users within the home to use the same chat environment to converse with each other directly and then easily direct their questions or actions to machine agents when needed.

It is the use of this common human-readable CE language that enables the passive observation of system state and agent communications at any time without development of special tooling to convert from machine specific representation formats to something that human users can directly read. The CE language enables machine or human users to change or extend the conceptual models the system is operating on, as well as allowing them to define new knowledge, queries or rules.

Whilst it would be possible to demonstrate the same capabilities using more traditional Semantic Web languages they would be aimed at machine processability rather than human consumability and would therefore require additional components to be developed to allow conversational interaction and the inclusion of the human users in the conversation.

IV. Evaluation

As set out in the introduction, our hypothesis is that CNL can enable machine-machine, machine-human, human-machine and human-human interaction in a dynamic environment. The previous section has given illustrative examples of how we envisage the approach working in a range of use cases.

A. *Earlier work*

Through a series of experiments, we are building an evidence base to show the feasibility and effectiveness of the approach, in two respects: (i) that humans without any significant degree of training are able to engage in dialogues using a combination of NL and CNL; and (ii) that the approach supports environments that can rapidly evolve to accommodate new devices and capabilities as they are encountered.

In earlier work we have sought evidence for (i), specifically: we have to date run a series of trials in controlled conditions, focusing on the proposition that users with little or no training in the use of CE can productively interact with CE-enabled agents. We reported the results of the first of these studies in [38]. Twenty participants (undergraduate students) were assigned a task of describing scenes depicted in photographs using NL, and given feedback in the form of CE statements generated via NLP by a software agent. The agent had been constructed rapidly to perform simple bag-of-words NLP with a lexicon provided by having four independent people provide scene descriptions in advance of the study. The results were promising; from 137 NL inputs submitted by the 20 subjects, with a median of one sentence for each input, a median of two CE elements was obtained by NLP for each input. In other words, with no prior training in the use of CE or prior knowledge of the domain model constructed for the scenes, users were able to communicate two usable CE elements (typically an identified instance and a relationship) per single-sentence NL input.

The ability of the CE agent to extract meaningful elements from the user's input and confirm these in CE form was constrained by the rapid construction of the background domain knowledge base. In effect, the agent's limited knowledge about the world led to results that were characterized by high precision, but relatively low recall, since the agent was engineered only to be "interested" in a narrow range of things. In this respect, however, we see these results as applicable to our Conversational Homes scenarios, where the concerns of home-based devices and the affordances users expect them to provide will be similarly narrow. Further studies are planned in settings more closely aligned with the examples in the previous section, and the remaining sections of this paper talk in more detail about the first specific Conversational Homes evaluation in this series.

In our second trial, 39 participants (undergraduate students) assigned to three groups conducted a crowdsourcing task using a conversational agent deployed on mobile devices, entering observations via NL and confirming machine-generated CE that was then added to a collective knowledge base in real time [43]. Usability of the conversational agent was operationalised as task performance [44]: the number of user-inputted NL messages that were both machine interpretable (i.e., could be mapped to CE) and confirmed by the user. Overall, despite close to no training, 74% of the participants inputted NL that was machine interpretable and addressed the assigned crowdsourcing task. Participants reported positive satisfaction based on scores from the System Usability Scale (SUS) [45], with means in the high 60s being consistent with good usability.

In terms of our requirement (ii), that the approach supports environments that can rapidly evolve to accommodate new devices and capabilities as they are encountered, we have constructed and demonstrated experimental prototypes for sensing asset selection for users' tasks, as described in [46]. Again, while these prototypes are not exactly aligned with the scenario of home automation (instead being more concerned with sensing systems such as autonomous aerial vehicles and ground systems) these experiments have shown that the CE-based approach supports the rapid addition of new knowledge. This includes not only of types of asset, but also of asset capabilities (that can be used to match assets to tasks). In many ways, the home setting is simpler than, say, an emergency response or search-and-rescue scenario, so we believe that the positive outcomes of these experiments are translatable into the domestic context.

An arguable difference between the home versus emergency response or search-and-rescue settings is the degree of training that a user can reasonably be expected to have obtained in the use of the available devices. In the home setting, this must always be minimal. In the other setting, however, minimal training is still desirable, since users should not necessarily be experts in the operation of sensing systems [47]. In any case, we argue that this usability point is addressed under (i) above. Also, in many cases, the addition of knowledge about new devices and their capabilities will typically be provided by the originators of the devices rather than end-users, though our approach does not preclude a "power" user from providing additional knowledge to their local environment.

B. User Evaluation

Based on the results summarised above that provided evidence that untrained users can quickly learn to interact with complex systems using our CNL and conversational technology, our most recent work aims to validate the Conversational Homes concept by means of a study with 12 participants. The primary goal of this research is to determine whether it is possible to build such an environment using a CNL basis as described earlier in this paper and, if so, whether the time and effort taken to do so is an improvement over traditional programming approaches. During the build phase factors such as time and complexity were not explicitly measured, although it should be noted that the entire model and fact-base for the evaluation were able to be successfully built entirely in the Controlled English language using the ce-store runtime environment without the need for any additional code or modifications to that environment. The entire end-to-end development time of the application was 2 days for 1 person, the majority being spent on developing the custom JavaScript code to render the live schematic view and the conversation. Within this 2 days development time only a couple of hours were spent on model and fact-base development, using just a plain text editor to write the CE languagestatements.

Since we are not trying to measure the comparative cost and benefit of the development time of environments such as these, the evaluation itself is therefore aimed at the untrained participants using the resulting environment. For this, we followed the same model as for our second trial summarised above, operationalising usabilty as task performance [43], specifically: whether participants can use the Conversational Homes chat interface to successfully interact with the environment to achieve simple goals or get simple information from the system as to the state of different components. The study was designed as a series of 5 simple tasks that were given to the participants in a group setting with each participant interacting with a separate local environment. The total available time for the study was 20 minutes, with each task taking 2-3 minutes. The tasks were communicated to the group via a shared projected screen with simple instructions; the instructions for each task (see subsections below) were shown to all users, with the text remaining on the screen for the duration shown. No additional information or guidance was given and the participants were free to use the conversational interface and/or the schematic to interact with the system as needed. Each task attempted to serve a different purpose, with the tasks, descriptions and planned purposes listed in the subsections below:

1) Task 1 – Simple query:

- Instructions: “Find out which lights are switched on...”
- Duration: 3 minutes
- Purpose: To establish whether the participants were able to use the conversational interface to determine the state of the lights. The live schematic view could also be used for this purpose since it shows the states of all lights, however we were expecting to see evidence of the participants asking this question via conversation.
- Success: The participant asks a question where the answer contains the state of all lights that are switched on.

2) Task 2 – Simple state change:

- Instructions: “Shut the front door”
- Duration: 2 minutes
- Purpose: State changes for items within the Conversational Homes can only be achieved via conversational interaction. To prevent the participants simply typing in the exact guidance text we deliberately did not specify the word “shut” in our model (e.g., as a synonym for “close”). This meant that participants must at least experiment with trivial restatements of the guidance in order to discover the correct term used to model the “close” action.
- Success: A statement from the participant that results in the “Front Door” state becoming “Closed”.

3) Task 3 – Group state change:

- Instructions: “Turn on all the lights in the bedroom”
- Duration: 2 minutes
- Purpose: Achieve a state change for a group of devices. We deliberately designed this task so that the users could type the exact guidance text into the system in order to achieve the desired result. This was to contrast with Task 2 where we deliberately left out the obvious form in order to force the user to seek alternatives.
- Success: One or more statements from the participant that results in all of the lights in the bedroom state becoming “On”.

4) Task 4 – Multiple state changes:

- Instructions: “It’s bedtime... Get the house in the right state for bed (It’s a hot night)”
- Duration: 3 minutes
- Purpose: The description for this task is deliberately ambiguous and more descriptive. This was intentional and designed to see whether the participants could successfully translate a generic and high-level desired state into specific actionable requests. We anticipated that this would be interpreted as switching off lights and opening the bedroom window.
- Success: Statements that result in (at least) the turning off of the bedroom lights and the opening or closing of the bedroom window. The suggestion that “It’s a hot night” was intended to elicit an opening of the window, however we realise that cultural differences could yield different responses, for example closing the window to ensure that air conditioning works efficiently. Even the turning off the bedroom lights may not meet every participants definition of the “...right state for bed...” but we needed to see some evidence of state change towards the target goal and therefore chose these two conditions as valid indicators of task completion.

5) Task 5 – Open ended:

- Instructions: "Freestyle: Talk to the house using phrases you'd actually want to use in real life. We have temperature sensors and door cameras plus we can add any other devices of capabilities into the system. They won't work but will be a great source of ideas :)"
- Duration: 5 mins
- Purpose: The purpose of this task was simply to elicit a wider range of interactions from the participants to inform the design of future evaluation studies and enable them to express themselves in ways that would be desirable to them in such a system.

The system was instrumented to record all conversational interactions (human and system generated) as well as all state changes that occurred. These interactions and state changes were subsequently analysed post-study for all users and are the basis for the results section (Section IV-E) later in this paper.

Note that the Conversational Homes environment is entirely simulated for this study: there is no linkage to actual sensors or devices in the physical world. It should be noted however that integrating this simulated environment into physical devices is extremely easy, assuming that the devices have APIs and are able to be queried and have their states changed programmatically. From an implementation perspective it is simply a case of recording the relevant information to locate and interact with each device (e.g., IP address, port and any required credentials) within the CE language. Having done so it is then trivial to write (in CE) a trigger that will be called each time a state change instance is generated by the system. The state change instance contains all details required to modify the device and the target state so the trigger in question can simply call some very generic code or invoke a generic web service that will simply invoke the target device API with the required credentials and parameters to make the desired change occur.

C. Participants

The participants for this study were drawn from a sample of convenience: they were all members of the IBM UK Emerging Technology team excluding authors of this paper and those familiar with the underlying research context and technologies being developed. The study was run once over a 30 minute period with the participants volunteering to participate in the study over their lunch break. There were a total of 12 participants, each of whom brought their own device (laptop or tablet) to enable them to participate in the study. All participants were "untrained users" with no prior experience of the technologies used in the study and had not previously seen or heard about the Conversational Homes user interface. Each participant was asked to login to the browser-based system and provide a unique userid (name) to be identified by. To ensure no capture of personal information, each of the usernames was substituted for a single letter in the range A-L post-study. In this paper users are thus referred to in this style (i.e., "User A") with this indicating a single anonymised human participant within the study. The entire cohort were located in a single large room with a projected screen displaying the guidance for each task. Participants were allowed to talk to each other if needed but we observed that generally the participants worked alone and with relatively little chatter or verbal discussion between them.

D. Design and Hypothesis

For this evalution, our hypothesis was that the Conversational Homes agent would have good usability, operationalised as performance of the five simulated household tasks.This builds upon our earlier general hypothesis that CNL can enable machine-machine, machine-human, human-machine and human-human interaction in a dynamic environment such as this, and accounts for the development of the specific conversational home agent and user interface.

The CE resources for this evaluation are publically available online [48] as is the custom browser-based user interface that was developed specifically for this evaluation [49]. Figure 5 shows this user interface and each participant within the study interacted solely via this environment. Each participant was operating entirely alone, with each conversational home being a separate instance purely for the use of that single participant and no ability for messages from one participant to access the contents or state of any other participants environment.

The user interface is broken into two fundamental components: The left-hand schematic view, and the right-hand chat interface.

The left-hand schematic view is a "live" representation of the conversational home that is built around a single level apartment unit. It is described as live since the state of all the sensors and actuators are rendered dynamically. This means that as lights are switched on or off, or as doors/windows are opened/closed, the visual state of the conversational home schematic is updated accordingly. These state changes happen regardless of the source of the change, i.e., if someone used some other interface to change the state then the schematic would update even if the change had not originated in the right-hand conversation pane.

The right-hand chat interface is built to mimic commonplace messaging applications that all users will already be familiar with on iOS and Android platforms. Messages from the human user are displayed in right-aligned green boxes and messages from the conversational home are displayed left-aligned in white boxes. There are no restrictions on the format, style or content of the messages that the human user can type, although the ability for the system to correctly interpret the messages from the human users is affected by brevity, precision and whether the text is "on-topic" for the conversational home environment.

There are a number of sensors, actuators and spaces that comprise the conversational homes environment and each of these are shown in the schematic in Figure 5. From bottom-left to top-right, the rooms and their contents are:

- Building Hallway
  - Front Door
- Front Room
  - Front-to-Hallway Door
  - Front Left Window
  - Front Right Window
  - Front Room Temp Sensor
  - Front Room Window Camera
  - Front Room Door Camera
  - Back Overhead (light)
  - Front Overhead (light)

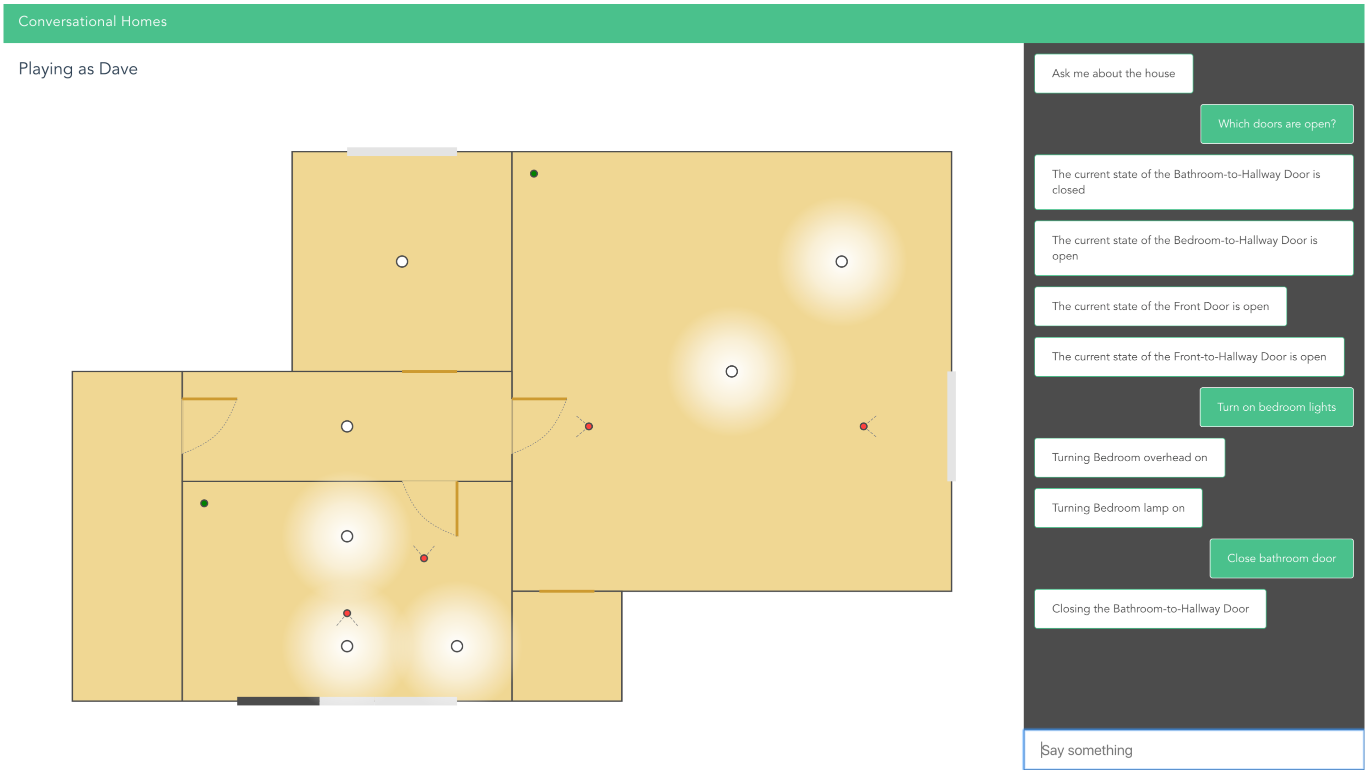

Figure 5: Conversational Homes User Interface.

 - Side lamp (light)
- Hallway
  - Hallway Overhead (light)
- Bathroom
  - Bathroom-to-Hallway Door
  - Bathroom Window
  - Bathroom Overhead (light)
- Cupboard
  - Cupboard-to-Bedroom Door
- Bedroom
  - Bedroom-to-Hallway Door
  - Bedroom Window
  - Bedroom Temp Sensor
  - Bedroom Door Camera
  - Bedroom Window Camera
  - Bedroom Overhead (light)
  - Bedroom Lap (light)

In Figure 5 The different states can be seen visually. For example: The Bathroom-to-Hallway Door is closed whereas the other Hallway doors are open, the Bathroom light is off whereas the Bedroom lights are on, and one of the Front room windows is open whereas the other windows are closed.

E. Results

The study was carried out on 19th August 2017 from approximately 12:30 to 12:50 British Summer Time. There were 12 participants, drawn from the IBM UK Emerging Technology team who sent a total of 367 messages across the 5 simulated household tasks.

The full set of results and corresponding guidance for each of the 5 tasks can be found online in the Open Science Framework [50]. The headline results were that our set of untrained users were able to quickly learn how to interact with the Conversational Homes system in order to find the state of various sensors and to interact with the environment to affect the correct state changes for the tasks. Of the 4 tasks undertaken with measurable success criteria: all users were able to successfully complete 3 out of 4 of the tasks and the fourth (more complex) task was successfully completed by 10 out of 12 of the users. On average each user was able to complete the 3 simple tasks in around 30 seconds, with the fourth (more complex) task taking around 1 minute 30 seconds on average to complete. The primary result, therefore was that the Conversational Homes agent had good usability in user performance of the simulated household tasks, consistent with our earlier experimental work reported previously.

Figure 6 shows the individual participants cumulative assertion counts during the study, with each of the time periods for the 5 tasks shown. We observe a steady upwards progressions for each user, even when taking into account only the valid assertions made during the study, i.e. those assertions deemed to have been correctly interpreted by the system. These results suggest that the proposed conversational approach mediated by the CNL technology can be effectively used with little or no training: a sizeable majority of users were able to make successful use of the interface and query or assert information in a relatively short period. They were largely able to complete their tasks in a short time period even when multiple attempts were required due to interpretation of their NL input was not initially successful.

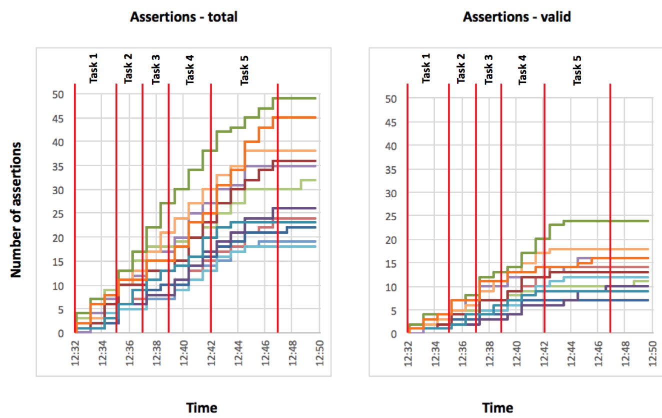

Figure 6: Individual Participants Cumulative Assertions over time.

TABLE I: Overall user statistics

User	Msgs sent	Msgs rcvd	State changes	Duration	Msg freq
A	19	91	39	13:11	42 secs
B	27	80	35	13:18	30 secs
C	34	74	49	16:51	30 secs
D	35	139	72	11:35	20 secs
E	19	91	41	11:59	38 secs
F	28	115	58	12:39	20 secs
G	22	50	8	15:01	41 secs
H	36	44	28	14:13	24 secs
I	51	128	72	18:20	22 secs
J	28	125	85	14:29	31 secs
K	24	131	31	11:04	28 secs
L	45	169	67	14:18	19 secs

Table I shows the overall user statistics for the study, showing the overall number of messages sent, received and the frequency of message sends for each user. The number of state changes are also shown; these are created each time a light is switched on or off, or a door/window is opened or closed. It is interesting to note that the message frequency is relatively low (in the range 20-41 seconds), suggesting that the participants were carefully considering their inputs rather than simply firing many messages in to the interface.

Table II shows the average number of NL messages and average time taken to complete each of the tasks. Note that Task 5 is not shown since it was a freestyle task with no completion criteria. The number of messages is low and the time to completion is fast for each of the tasks with the exception of task 4. Task 4 required multiple separate messages in order to complete and was deliberately ambiguous in order to better test the participants and this can be seen in the resulting time-to-completion statistics.

These results show that the Conversational Homes chat interface could be plausibly useful to untrained novice users

TABLE II: Task level statistics

Task	Msgs to complete	Time to complete	Completion
Task 1	1.3	00:33	12
Task 2	1.9	00:35	12
Task 3	1.3	00:26	12
Task 4	4.2	01:31	10

even with the relatively immature level of NL processing currently available within our CNL-based solution. Even with a relatively high level of failed interpretations the participants were generally able to achieve the stated task goals in a short period of time and with relatively few messages. We believe that this is an encouraging result and justifies the pursuit of further evidence in the future to determine whether other factors like the richness of the model contribute to observable outcomes such as the ability to complete tasks or the number of misinterpretations of the NL messages.

As in our earlier usability study, subjective usability assessment was performed by asking the participants to complete the SUS [45] questionnaire. Participants reported positive satisfaction with a mean of 69 indicating a good degree of perceived usability across the group, consistent with the previous results in [43] and providing converging evidence for the usability of the conversational agent.

F. *Observations and Discussion*

In order to process the results we manually reviewed each of the sent messages to determine whether the system had correctly interpreted the text, and in cases of misinterpretation what category of misinterpretation it was.

1) Text interpretation analysis: Each of the text messages has been analysed and classified as “successfully interpreted” or not, however there are a number of common reasons for incorrect interpretation as well as some potential improvements to even the successfully interpreted messages.

In all of the cases of misinterpretation, or interpretation improvements, there are simple remedial actions that can be taken in the CE model and fact-base to catch each of these cases and handle them correctly. This is often through the addition of synonyms or through extension of the model to support additional capabilities not envisaged in the original modelling exercise (e.g., the addition of new device types, or new actions for existing device types). In all of these cases the effort required to extend the model is small, and the level of technical skill is low. The tooling (such as ce-store) that has been developed for studies in environments such as these mean that the changes can be deployed extremely quickly, potentially in real-time as issues are identified and resolved. However this entire approach is based on the assumption that all possible phrases and terminology could be identified in advance and therefore designed into the model and fact-base. This leads to a system with a high degree of accuracy and a low false-positive rate but one that is very brittle and must be focused on a particular domain to achieve that accuracy. Better hybrid approaches may be possible, for example using Machine Learning techniques to classify messages into a form that can be handled by the underlying CE models. This would be a small change but may enable a much wider set of phrases to be correctly handled with the existing system and without needing to continually update the CE model as new language is encountered. This approach could also help to handle evolving terminology and slang as it is adopted by the user community.

The remaining sub-sections deal with each of the types of misinterpretation encountered during this study.

2) Successful interpretations with room for improvement: These messages are classified as correctly interpreted since they do give enough information to provide the answer requested, or perform the action requested. However, from a human interpretation perspective they would be perceived to be "not quite right", usually due to some violation of a Gricean maxim [42] such as quantity:

- Too much information
  e.g., *"Which lights are on?"*
  Currently lists the states of all lights, including those that are off. This should really only list lights in the state specified.
- Ignoring unknown qualifiers
  e.g., *"Turn all the lights off downstairs"*
  The action was performed globally, e.g., turning off all the lights, and the unknown qualifier (e.g., downstairs) was quietly ignored. A more human-like reaction would be to either question the unknown qualifier or at least confirm that it was ignored.
- Ignoring current state
  e.g., *"Turn on bedroom lights"*
  The action was correctly interpreted and carried out regardless of the current state of the lights. A more human response would be to feedback that the lights were already on rather than saying they were switched on in cases where they were already on.
- Common sense defaults
  e.g., *"Close the bedroom door"*
  This would result in the Bedroom-to-Hallway door being closed and the Cupboard-to-Bedroom door being closed. This is arguably the incorrect behavior as the user probably meant the "main" bedroom door (Bedroom-to-Hallway door) rather than all bedroom doors.

3) Misinterpretations: In many cases the messages sent by the human users were not able to be interpreted at all. The reasons for these misinterpretations fall into a number of categories described below:

- Spelling mistakes
  e.g., *"turn on th elights"*
  Using our CNL and defined vocabulary based approach it is simply impossible to handle all possible misspellings and typos. As mentioned previously, this is an area where the use of additional lexical analytics or machine learning capabilities could be useful to augment the basic CNL solution to better handle issues such as typos.
- Unmapped synonyms
  e.g., *"shut"*
  We deliberately left out "shut" as a synonym for "close" to see whether participants would attempt obvious alternatives. Other less obvious synonyms were simply missed due to the speed of implementation and lack of testing.
- Split phrases
  e.g., *''Turn all the lights on"*
  Since "Turn on" is the trigger phrase this common practice of splitting the phrase "turn on" with the subject ("all the lights"), i.e., "Turn something on", causes the trigger phrase to be ignored. This is easily addressed through a simple additional lexical extension to the natural language processing capability within ce-store to handle split phrases such as these.
- Ignoring key filter words
  e.g., *"Only list the lights that are on"*
  Since all lights are listed regardless of state this means the message was not interpreted correctly as the intent was clearly to filter using "only" to list those of a specified state.
- Ignoring key action words
  e.g., *"How many lights are on"*
  This should return a number rather than a list of all lights and their individual states.
  e.g., *"What is x?"* or *"Where is x?"*
  This should return a description (or location) of the item in question.
- Conditionals
  e.g., *"if something then do something"*
  Some users indicated these kinds of rules, which were often interpreted by performing the action and ignoring the conditional.
- Out of scope
  e.g., *"Turn on all the cameras"*, or *"lock the doors"*, or *"What is the temperature in the bedroom"*, or *"Turn up the temperature"*
  A number of sensors were modeled and shown on the schematic but unavailable for interaction. Some users wrote messages to attempt to interact with these items.
- Off topic
  e.g., *"Import velociraptors"*, or *"activate discoball"*
  Messages relating to things that were not defined in the model. These mainly came in Task 5, which was the open-ended "Freestyle" task designed to elicit open-ended and off topic messages.
- Complex sentences
  e.g., *"Which rooms are the lights switched on in?"*
  The response should give a list of rooms, not a list of lights.

G. Limitations

In Section III we describe the extent of our research into CNL technology and the full conversational protocol that can be supported based on our earlier work to model speech act theory. Evaluation of this full protocol is not possible within this initial study and the conversations that are possible between human users and machine agents are limit to simple single-turn "tell" and "ask/tell" interactions with responses coming back to the human user in gist forms. By single-turn we mean that there is no possibility for reference back to previous statements within the current study. This rules out styles of interaction such as anaphoric reference (e.g., "Are the lights in the bedroom on?", followed by "Ok, turn them on please" where "them" is an anaphoric reference to "the lights in the bedroom" from the previous dialogue phrase). Again, the research basis for this work does explicitly support multi-turn dialogue and features such as anaphoric reference but these were not enabled for this study.

Another key design decision was that within this particular study the human users are never exposed to the raw CNL of the underlying system. Specifically, in Section III-B we discuss two possible modes of interaction between the human users and the system components: "Concierge only" interactions, and "Direct communication to any device" interactions (including an optional concierge agent if needed). We also note that the latter is the most powerful of the two interaction patterns and that in the latter cases all devices and human users would be able to interact via CNL. For this initial study we have implemented the "concierge only" mode as the is the initial exercise in a planned series of experiments intended to further develop the capabilities of the system and the devices within the simulated environment. Communication between the human users and the system is always via a single concierge agent, and in cases where state changes are needed (e.g., switching on a light) the human user "tells" the concierge in NL the desired state change, the concierge agent attempts to interpret the NL into CNL and passes the CNL into the ce-store to reflect the state change requested by the user. Within our conversational protocol it is possible for the concierge to show the CNL to the human user to seek their confirmation (in the form of a "confirm" card) and only pass confirmed CNL into ce-store, but for this study we felt that this confirm stage was not needed due to the simplicity of the tasks being undertaken. In the results analysis we do identify cases (especially "Ignoring unknown qualifiers") where a confirm step back to the human user would help in some cases to prevent unexpected outcomes.

Finally, we note that of the 4 interaction styles listed in Section III-D we only support two in this study. Support for the other two (rationale and implicit desire to change state) are more subtle and advanced cases and may be considered for future studies. The two supported interaction styles are "Direct question/answer exchange" and "An explicit request to change a particular state".

V. Conclusion and Future Work

In this paper we have explored the use of conversations between humans and machines, motivated by a desire for "beautiful seams". We assert that this approach could enable better understanding of complex system such as a set of IoT devices in a home. In this paper, we have shown how semantic representations can be used in a human-friendly format through the use of a CNL technology known as ITA CE. Through the use of a conversational protocol built on top of the core CE language we show how human and machine agents are able to communicate using this single language. Examples of the CE language are provided throughout the paper showing how different concepts can be constructed and the subsequent data for the knowledge base can be provided in the same CE language. Through a set of four typical types of interaction we show how human users can interact with the devices in such an environment, and we note that whilst we have focused these four examples on a human-machine interaction, the exact same approach applies to machine-machine as well. Some additional discussion around what machine-human and human-human forms would look like is mentioned. Building on the initial success with the study reported in Section IV future work may include designing and conducting more advanced experiments in the conversational home setting, specifically to address some of the limitations described in Section IV-G and advance our experimental capability to better reflect the full potential identified in the CNL and conversational research work.

Our work also continues into the wider investigation into the potential for human-machine conversational capabilities, specifically with the JavaScript-based CENode component [40], which is designed specifically for CNL processing at the very edge of the network (directly in the end users mobile phone or tablet browser environment). Some of this latest work is informally reported for IoT interactions [51] and integration with the popular Alexa platform [52]. In our desire to investigate the potential for integration of machine learning capabilities into the core CNL approach we also plan to investigate easy to use online services such as IBM Watson Conversation [53] as a potential front-end and dialogue orchestration component, which would process all incoming NL from the human users and convert into a simplified form before presenting to our CNL implementation. If successful this could provide a significant improvement in handling a wider set of synonyms, spelling mistakes and other forms of evolving language without needing to predefine them in the CNL environment.

Acknowledgment

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Numbers W911NF-06-3-0001 and W911NF-16-3-0001. 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 U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] N. OLeary, D. Braines, A. Preece, and W. Webberley, "Conversational homes," in *9th International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE17)*, 2017, pp. 82–89.

[2] R. Dale, "The return of the chatbots," *Natural Language Engineering*, vol. 22, no. 5, pp. 811–817, 2016.

[3] M. Witbrock and L. Bradeško, "Conversational computation," in *Handbook of Human Computation*. Springer, 2013, pp. 531–543.

[4] J. Lester, K. Branting, and B. Mott, "Conversational agents," *The Practical Handbook of Internet Computing*, pp. 220–240, 2004.

[5] N. O'Leary. (2014) Conversational iot. [Online]. Available: http://knolleary.net/2014/12/04/a-conversational-internet-of-things-thingmonk-talk/ (Visited on 27-Nov-2017)

[6] A. Stanford-Clark. (2008) Redjets on twitter. [Online]. Available: https://twitter.com/redjets (Visited on 27-Nov-2017)

[7] T. Armitage. (2008) Tower bridge on twitter. [Online]. Available: https://twitter.com/twrbrdg_itself (Visited on 27-Nov-2017)

[8] (2008) Mars curiosity on twitter. [Online]. Available: https://twitter.com/MarsCuriosity (Visited on 27-Nov-2017)

[9] (2010) Philae lander on twitter. [Online]. Available: https://twitter.com/Philae2014 (Visited on 27-Nov-2017)

[10] (2011) Rosetta probe on twitter. [Online]. Available: https://twitter.com/ESA_Rosetta (Visited on 27-Nov-2017)

[11] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Communications of the ACM*, vol. 59, no. 7, pp. 96–104, 2016.

[12] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[13] M. Weisser, "The computer for the twenty-first century," *Scientific American*, vol. 265, no. 3, pp. 94–104, 1991.

[14] M. Chalmers, "Seamful design and ubicomp infrastructure," in *Proceedings of Ubicomp 2003 Workshop at the Crossroads: The Interaction of HCI and Systems Issues in Ubicomp*. Citeseer, 2003.

[15] T. Coates. (2014) Interacting with a world of connected objects. [Online]. Available: https://medium.com/product-club/interacting-with-a-world-of-connected-objects-875b4a099099#.nd00bbs5n (Visited on 27-Nov-2017)

[16] Hypercat. [Online]. Available: http://hypercat.io (Visited on 27-Nov-2017)

[17] M. Bates and R. M. Weischedel, *Challenges in natural language processing*. Cambridge University Press, 2006.

[18] R. L. C. Delgado and M. Araki, *Spoken, multilingual and multimodal dialogue systems: development and assessment*. John Wiley & Sons, 2007.

[19] B. Dumas, D. Lalanne, and S. Oviatt, "Multimodal interfaces: A survey of principles, models and frameworks," *Human machine interaction*, pp. 3–26, 2009.

[20] G. Churcher, E. S. Atwell, and C. Souter, *Dialogue management systems: a survey and overview*. University of Leeds, School of Computing Research Report 1997.06. 1997., 1997.

[21] P.-H. Su, M. Gasic, N. Mrksic, L. Rojas-Barahona, S. Ultes, D. Vandyke, T.-H. Wen, and S. Young, "Continuously learning neural dialogue management," *arXiv preprint arXiv:1606.02689*, 2016.

[22] D. R. Traum and S. Larsson, "The information state approach to dialogue management," in *Current and new directions in discourse and dialogue*. Springer, 2003, pp. 325–353.

[23] D. Mouromtsev, L. Kovriguina, Y. Emelyanov, D. Pavlov, and A. Shipilo, "From spoken language to ontology-driven dialogue management," in *International Conference on Text, Speech, and Dialogue*. Springer, 2015, pp. 542–550.

[24] M. S. Yakoub, S.-A. Selouani, and R. Nkambou, "Mobile spoken dialogue system using parser dependencies and ontology," *International Journal of Speech Technology*, vol. 18, no. 3, pp. 449–457, 2015.

[25] G. Meditskos, E. Kontopoulos, S. Vrochidis, and I. Kompatsiaris, "Ontology-driven context interpretation and conflict resolution for dialogue-based home care assistance," in *Paschke A, Burger Al, Splendiani A, Marshall MS, Romano P. Proceedings of the 9th International Conference Semantic Web Applications and Tools for Life Sciences (SWAT4LS); 2016 Dec 5-8; Amsterdam, The Netherlands.[Place unknown]:[CEUR]; 2016.[5 p.]*. CEUR Workshop Proceedings (CEUR-WS. org), 2016.

[26] T. Kuhn, "A survey and classification of controlled natural languages," *Computational Linguistics*, vol. 40, no. 1, pp. 121–170, 2014.

[27] R. Schwitter, "Controlled natural languages for knowledge representation," in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 2010, pp. 1113–1121.

[28] D. Mott, "Summary of ita controlled english," *ITA Technical Paper, http://nis-ita.org/science-library/paper/doc-1411a (Visited on 27-Nov-2017)*, 2010.

[29] A. Preece and W. R. Sieck, "The international technology alliance in network and information sciences," *IEEE Intelligent Systems*, vol. 22, no. 5, 2007.

[30] D. Mott, C. Giammanco, M. C. Dorneich, J. Patel, and D. Braines, "Hybrid rationale and controlled natural language for shared understanding," *Proc. 6th Knowledge Systems for Coalition Operations*, 2010.

[31] T. Klapiscak, J. Ibbotson, D. Mott, D. Braines, and J. Patel, "An interoperable framework for distributed coalition planning: The collaborative planning model," *Proc. 7th Knowledge Systems for Coalition Operations*, 2012.

[32] D. Braines, D. Mott, S. Laws, G. de Mel, and T. Pham, "Controlled english to facilitate human/machine analytical processing," *SPIE Defense, Security, and Sensing*, pp. 875 808–875 808, 2013.

[33] J. Ibbotson, D. Braines, D. Mott, S. Arunkumar, and M. Srivatsa, "Documenting provenance with a controlled natural language," in *Annual Conference of the International Technology Alliance (ACITA)*, 2012.

[34] F. Cerutti, D. Mott, D. Braines, T. J. Norman, N. Oren, and S. Pipes, "Reasoning under uncertainty in controlled english: an argumentation-based perspective," *AFM*, 2014.

[35] D. Braines, J. Ibbotson, D. Shaw, and A. Preece, "Building a living database for human-machine intelligence analysis," in *Information Fusion (Fusion), 2015 18th International Conference on*. IEEE, 2015, pp. 1977–1984.

[36] D. Mott and J. Hendler, "Layered controlled natural languages," in *3rd Annual Conference of the International Technology Alliance (ACITA)*, 2009.

[37] A. Preece, D. Braines, D. Pizzocaro, and C. Parizas, "Human-machine conversations to support multi-agency missions," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 18, no. 1, pp. 75–84, 2014.

[38] A. Preece, C. Gwilliams, C. Parizas, D. Pizzocaro, J. Z. Bakdash, and D. Braines, "Conversational sensing," in *SPIE Sensing Technology+ Applications*. International Society for Optics and Photonics, 2014, pp. 91 220I–91 220I.

[39] D. Braines. (2015) Ita controlled english store (ce-store). [Online]. Available: https://github.com/ce-store (Visited on 27-Nov-2017)

[40] W. Webberley. (2016) Cenode.js. [Online]. Available: http://cenode.io/ (Visited on 27-Nov-2017)

[41] If this then that. [Online]. Available: https://ifttt.com/ (Visited on 27-Nov-2017)

[42] H. P. Grice, "Logic and conversation," *1975*, pp. 41–58, 1975.

[43] A. Preece, W. Webberley, D. Braines, E. G. Zaroukian, and J. Z. Bakdash, "SHERLOCK: Experimental evaluation of a conversational agent for mobile information tasks," *IEEE Transactions on Human-Machine Systems*, vol. in press, 2017.

[44] J. Nielsen, *Usability engineering*. AP Professional, 1994.

[45] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[46] A. Preece, D. Pizzocaro, D. Braines, and D. Mott, "Tasking and sharing sensing assets using controlled natural language," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 838 905–838 905.

[47] A. Preece, T. Norman, G. de Mel, D. Pizzocaro, M. Sensoy, and T. Pham, "Agilely assigning sensing assets to mission tasks in a coalition context," *IEEE Intelligent Systems*, vol. 28, no. 1, pp. 57–63, 2013.

[48] Github - conversational homes controlled english. [Online]. Available: https://github.com/ce-store/conv-homes (Visited on 27-Nov-2017)

[49] Github - conversational homes user interface. [Online]. Available: https://github.com/annaet/conversational-homes-viz (Visited on 27-Nov-2017)

[50] Conversational homes - experimental results. [Online]. Available: https://osf.io/pfskx/ (Visited on 27-Nov-2017)

[51] Cenode in iot. [Online]. Available: https://flyingsparx.net/2017/06/26/cenode-iot/index.html (Visited on 27-Nov-2017)

[52] Alexa, ask sherlock. [Online]. Available: https://flyingsparx.net/2017/07/19/cenode-alexa/index.html (Visited on 27-Nov-2017)

[53] Ibm watson conversation service. [Online]. Available: https://www.ibm.com/watson/services/conversation/ (Visited on 27-Nov-2017)