Controlled Natural Language

Dave Braines (IBM UK)
Controlled English as a pervasive information representation format could underpin a productive human/machine information processing environment.

What is Controlled Natural Language?

- A structured sub-set of natural human language
- Designed to reduce ambiguity and complexity
- Enable direct machine processability of human language information

Controlled English (CE)

- Our specific implementation of a Controlled Natural Language
- Based on earlier work by John Sowa
- Universal syntax and broad coverage
- Research focus of ITA and multiple transition projects
- Compatible with existing information representation & query standards

Insight: "CE store" – an instantiation of Controlled English in a multi-user dynamic information processing environment.
What can you do with Controlled English?

- Extensible model
- Universal syntax
  - Model – *conceptualise the person P1 is married to the person P2.*
  - Facts – *the person Fred is married to the person Jane.*
  - Queries – *for which P1 is it true that (the person P1 is an employee) and…*
  - Rules – *if the person P1 is married to the person P2 then the person P1 is a married person.*
  - Rationale - … *because the intel source UKI confirmed the sighting s0053.*
  - …and more

- Rich expressivity
  - Based on formal logic
  - Assumptions, truth values, uncertainty, trust levels etc
  - Compatible with open standards (OWL, RIF etc)

- Support for layers and extensions
  - More convenient ways of saying thing:
    - *the person Fred is described as “suspiciously walking into the airport”.*
So, why use Controlled English?

- Increased agility and effectiveness
  - Especially in new, dynamic, evolving situations
- Directly enabling non-technical users
- Improving human/machine hybrid capability
- A pervasive human-friendly representation
  - Encourage small-scale human updates – “local data from the edge”
  - Good basis for interacting agents (and humans)
- Increased confidence and transparent accountability
  - Through rationale
- Extensibility and universal syntax
- Opportunities for human language translation support
  - Plus alternative languages like Controlled French, German, Spanish etc

Numerous research collaborations, demonstrations and transition projects are experimenting with CE related capabilities both inside and outside the ITA programme.
Why are we researching this?

- Bringing capabilities of machines and humans together
  - Machine processing for some tasks
  - Human brain for insight & understanding
  - Machine processing to guide the construction of a valid rationale

- The key component is the human
  - Harnessing “collective intelligence”
  - Helping to make connections outside the system

- For analysts: How can we make them more effective and reduce their workload?
  - What is it “safe” to let the machine do?
  - The analyst must always be in control
  - What is the cost to the analyst?

The potential value of a “usable” semantic processing environment in the hands of non-technical, domain specialist, users is very high, especially in dynamic situations.

Existing Semantic standards offer great promise but are hampered by the complexity of the languages and required components. CE is a directly compatible human-friendly alternative with significant additional capabilities.
A set of examples from on-going research and transition:

- **Data-to-decision**
  Processing short geo-located messages from parties with mobile devices at the edge of the network

- **Information extraction**
  Processing natural language sources to extract intelligence information

- **Sensor to mission matching**
  Supporting the “CE-SAM” environment

- **UK open data**
  Processing of various UK open data sources

- **LOSA / fabric integration**
  Showing agility when consuming fast-flowing structured messages

- **Open source information processing**
  Newsfeed, twitter etc

All built with the CE store.
Providing an experimental Controlled English based information environment