Centre for
Data Analytics

Insight

A Brief State of the Art of CNLs for Ontology Authoring

Hazem Safwat and Brian Davis

DCU

Ollscoil na Gaillimhe
NUI Galway

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

UCD
DUBLIN

SFI
Science Foundation Ireland

Outline

- Introduction
- Main CNLs for the Semantic Web
  - Attempto Controlled English (ACE) & ACE Tools
  - Grammatical Framework (GF) & ACEWiki-GF
- Other CNLs
  - RABBIT
  - RABBIT to OWL Ontology authoring (ROO)
- Generation Driven CNLs
  - WYSIWYM
  - Round Trip Ontology Authoring (ROA)
- Evaluation of CNLs
- Conclusion

Motivation

- Provide a quick introduction about the state-of-the-art ontology authoring CNLs for other researchers.
- Recommend the most suitable CNL for a specific application in the semantic web, according to their evaluation with other CNLs.
- A potential for journal that includes all up to date semantic web CNLs and their respective evaluations and applications (in progress).
- The CNLs evaluation will be based on a 1 to 5 scale, for the quality of the evaluation, comparison with other CNLs, and the availability of statistical results.

Introduction

CNLs are defined as “subsets of natural language whose grammars and dictionaries have been restricted in order to reduce or eliminate both ambiguity and complexity”.

The need for Formal data representation and ontology creation and subsequently adopting semantic technologies challenges researchers to develop user-friendly means for ontology authoring.

CNLs for knowledge creation and management offer an attractive alternative for non-expert users wishing to develop small to medium sized ontologies.

Evaluation Quality Scale

1. No evaluation was performed
2. General evaluation through a pilot study (students only)
3. Good test subjects may or may not have statistical results but no comparison was held
4. comparative evaluation with other CNLs was held without statistical evidence
5. comparative evaluation with other CNLs plus statistical evidence

Main CNLs for the Semantic Web

Attempto Controlled English (ACE)

It is a subset of standard English designed for knowledge representation and technical specifications, and constrained to be unambiguously machine readable

CNL	Author/Year	Input	Output languages	Parsing Technique	User Evaluation	Multilingual	Category
Attempto Controlled English (ACE)	Fuchs, Kaljurand, and Kuhn (2008)	ACE Texts	Discourse Representation Structures (DRSs), which are syntactical variants of first order logic	Attempto Parsing Engine (APE) consists of a definite clause grammar written in Prolog.	5 ACE Vs. Manchester [1].	ACE-in-GF	CNL

ACEView

A plugin for the Protégé editor. It empowers Protégé with additional interfaces based on the ACE CNL in order to create, browse and edit an ontology.

CNL	Author/Year	Input	Output languages	Parsing Technique	User Evaluation	Multilingual	Category
ACEView	K. Kaljurand (2008)	ACE Texts	Discourse Representation Structures (DRSs	Attempto Parsing Engine (APE)	5 Rabbit OWL Vs. AceView [2]	-	Ontology Editor

ACEWiki

A monolingual CNL based semantic wiki that takes advantage of codeco ACE editor for its syntactically user friendly formal language, and it can translate all AceWiki content to OWL.

CNL	Author/Year	Input	Output languages	Parsing Technique	User Evaluation	Multilingual	Category
ACE Wiki	T. Kuhn (2008)	ACE Texts	Semantic Web Language (OWL)	ACE Parser (APE)	3 [3]	-	Ace Editor for wikis

Grammatical Framework (GF)

A framework for multilingual grammar , Rule based, Bidirectional.

Task : Machine Translation of controlled natural languages

The GF libraries now contain grammars for around 20 natural languages

CNL	Author/Year	Input	Output languages	Parsing Technique	User Evaluation	Multilingual	Category
GF	A. Ranta (2004)	Grammar Rules	CNL	Functional programming Language based on Haskell mapping from strings to trees	-	Yes	programming language for implementing CNLs

ACEWIKI-GF

A multilingual extension of AceWiki in addition to the multilingual environment.

modifying the original AceWiki to include GF multilingual Ace grammar, GF parser, GF source editor, and GF abstract tree

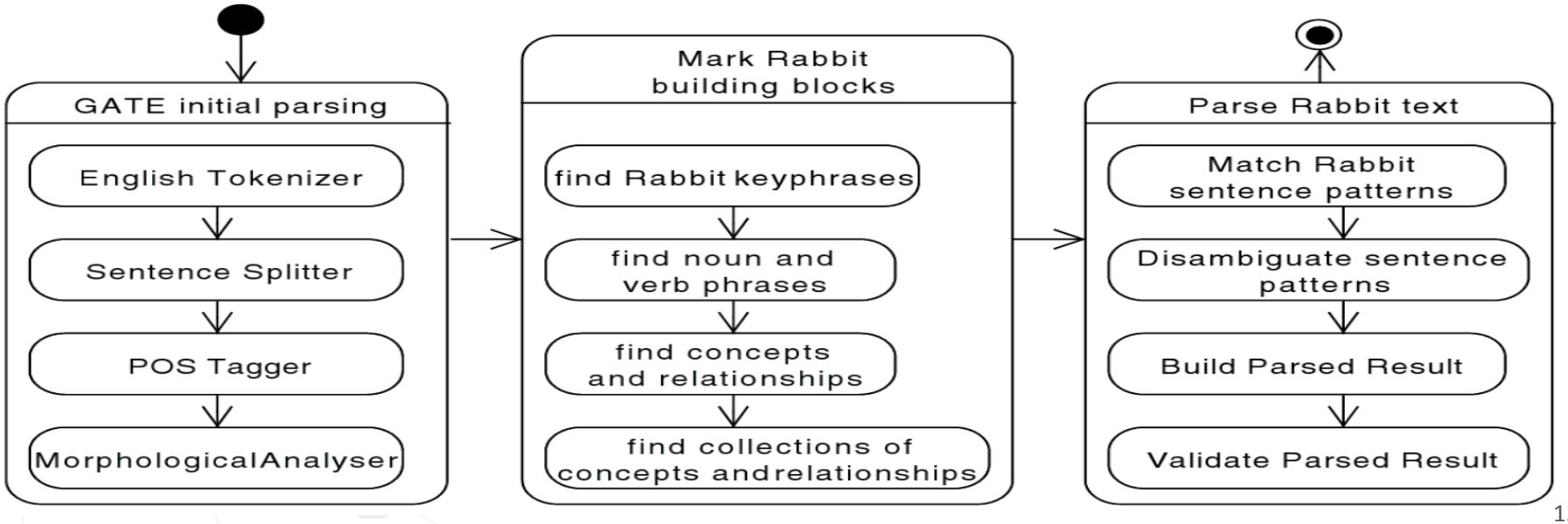
CNL	Author/Year	Input	Output languages	Parsing Technique	User Evaluation	Multilingual	Category
ACEwiki GF	K.Kaljurand & T. Kuhn (2013)	ACE Texts	Semantic Web Language (OWL)	ACE Parser (APE) + GF	5 AceWiki Vs. ACEWiki-GF [4]	Yes	Multilingual ACEWIKI

Other CNLs for the Semantic Web

RABBIT [5]

Developed and used by Ordnance Survey, Great Britain's national mapping agency, for the communication between domain experts and ontology engineers to create ontologies

Extension of Controlled Language for Ontology Editing CLOnE, implemented using the GATE framework.



RABBIT

CNL	Author/Year	Input	Output	Parsing Technique	User Evaluation	Multilingual	Category
Rabbit	Hart et al. (2008)	Rabbit Sentences	Parse Tree that gives access to the information found by the processing resources	Rabbit parser implemented in java consists of a pipeline of linguistic processing resources 1. preprocessing 2.Gazzetter 3.JAPE transducers	2. pilot study In [6] . 3. test subjects with no comparison [7]	Semantic Media wiki Yayan (a.k.a. Rabbit Chinese) CNL generator (CNLG) was implemented [8]	CNL

Rabbit to OWL Ontology authoring (ROO)

Ontology editing tool developed by the University of Leeds and is an open source Java based plug-in for Protégé.

supports the domain expert in creating and editing ontologies using Rabbit.

CNL	Author/Year	Input	Output languages	Parsing Technique	User Evaluation	Multilingual	Category
ROO	V. Dimitrova et al. (2008)	Rabbit Sentences	OWL	OWL API	5. OWL Vs. AceView [9]	-	Tool

Generation Driven CNLs

What you see is what you meant - WYSIWYM

Ontology editing tool, can edit a knowledge based reliably by interacting with natural language menu choices and the subsequently generated natural language feedback which can then be extended or re-edited using the menu options.

CNL	Author/Year	Input	Output languages	Parsing Technique	User Evaluation	Multilingual	Category
WYSIWYM	R.Power et. al. (1998)	Natural Language	Semantic Web Language (OWL)	WYSIWYM Engine built in Prolog + User interface in JAVA Natural Language Generator to provide feedback for user intractions	3. [10]	Editing Any language	Ontology Editor

Round Trip Ontology Authoring (ROA)

- builds on and extends the existing advantages of the CLOnE software.

It generates the entire CNL document first using SimpleNLG that is less sophisticated than WYSIWYM.

CNL	Author/Year	Input	Output languages	Parsing Technique	User Evaluation	Multilingual	Category
ROA	Davis et al. (2008)	Natural Language	Semantic Web Language (OWL)	GATE provides the JAPE (Java Annotation Pattern Engine)	5. ROA Vs. Protégé [11]	-	Ontology Editor

Evaluation of CNLs

ACE Evaluation

Ontographs are a graphical notation to enable tool independent and reliable evaluation of the human understanding of a given knowledge representation language.

The author categorizes CNLs evaluations into (1) task-based, whereby users are provided with a specific task to complete, and (2) paraphrase-based which are concerned with testing the understandability of the CNL.

The experiments compared the syntax of the ACE versus OWL framework called simplified Manchester OWL to test which framework is better in terms of, understandability, learning time, and users acceptance.

The results showed that users were able to do better classification using ACE with approximately 5% more accuracy than Manchester OWL, and 4.7 minutes less for learning and testing. Also, in terms of understandability ACE got a higher score than Manchester OWL.

Rabbit Evaluation

A paraphrase-based evaluation to assess whether domain experts without ontology authoring development can author and understand declaration sentences in Rabbit.

51% of the sentences generated at least one error. The most common error was the omission of the quantifier at the beginning of every sentence.

Another pilot study interested in whether people with no prior knowledge about Rabbit and no training in computer science, would be able to understand and correctly interpret and author sentences in Rabbit.

13 of the sentences were answered correctly by 75% of participants. These sentences were deemed sufficiently understandable by most participants.

ROO Evaluation

An evaluation study of ROO was conducted against ACEView where participants from the domains of geography and environmental studies were asked to create ontologies based on hydrology and environmental models, respectively

Although ACEView users were more productive, they tended to create more errors in the resulting ontologies. ROO users, their understanding of ontology modelling improves significantly in comparison to ACEView

Also, none of the ontologies produced were usable without post editing.

With respect to the extension of ROO, the study showed that 91% of the feedback messages were helpful to the users, and 78% were informative. However, feedback caused confusion and overwhelming for 10% of the cases.

WYSIWYM Evaluation

An evaluation of WYSIWYM was carried out with 16 researchers and PhD students from the social sciences domain.

The goal was to reproduce the descriptions using the WYSIWYM tool.

users mean completion times decreased significantly. Hence, users gained speed over time.

user feedback was positive, however the results were less positive in comparison to an earlier evaluation of WYSIWYM whereby users completion of tasks was less accurate.

Conclusions

Research within the CNL community is turning its attention towards multilingual controlled languages, with recent efforts to generate ACE, using GF, for several European languages.

There has been an increasing tendency towards conducting proper user evaluation for CNLs. While some CNL researchers have conducted task based evaluations, there have been less comparative evaluations across tools.

In general, the CNL community should invest more in conducting strong user evaluations and not to lose track of the end goal - the creation of more user friendly ontology editing interfaces.

Conclusions

A major question is whether a CNL is appropriate for the task?

there should be a pre-existing use case for a human orientated CNL, in other words a restricted vocabulary or syntax for a technical domain either legal, clinical or aeronautics such as ASD Simplified Technical English. Without such a use case there would be little incentive for users interact with it.

Factors to be taken into account when designing CNLs:

- Knowledge creation task complexity
- Target user (specialist or non expert)
- The domain (open or specific),
- Available corpora
- Sample texts
- Language resources or vocabularies, ontologies, multilingualism
- Requirements for language generation capabilities
- Availability of an NLP engineer or computational linguist

References

[1] T. Kuhn, **The understandability of OWL statements in controlled English Semantic Web**, Vol. 4, No. 1, 1 January 2013, pp. 101-115.
[2] Kaarel Kaljurand. **ACE View — an ontology and rule editor based on controlled English**. *Proceedings of the Poster and Demonstration Session at the 7th International Semantic Web Conference (ISWC2008)*, CEUR Workshop Proceedings, 2008.
[3] Tobias Kuhn**. AceWiki: A Natural and Expressive Semantic Wiki**. In Semantic Web User Interaction at CHI 2008: Exploring HCI Challenges, 2008.
[4] Canedo et al.**Evaluations of ACE-in-GF and of AceWiki-GF** (2013).
[5] Denaux et. Al. **Rabbit to OWL:Ontology Authoring with a CNL Based Tool with a CNL based tool**. CNL (2009), LNCS, vol.5972, pp.246-264. Springer, Heidelberg (2010).
[6] Hart, G., Johnson, M., Dolbear, C.: **Rabbit: Developing a control natural language for authoring ontologies**. In: Proceedings of the 5th European Semantic Web Conference. pp. 348–360. Springer (2008).
[7] Paula C. Engelbrecht, Glen Hart, Catherine Dolbear: **Talking Rabbit: A User Evaluation of Sentence Production**. CNL 2009:56-64

References

[8] Jie Bao et. Al. **A Controlled Natural Language Interface for Semantic MediaWiki**. Annual conference of the International Technology Alliance (2009).

[9] Vania Dimitrova, Ronald Denaux, Glen Hart, Catherine Dolbear, Ian Holt, and Anthony G.Cohn. **Involving Domain Experts in Authoring OWL Ontologies**. In Proceedings of the 7th International Semantic Web Conference (ISWC 2008), volume 5318 of Lecture Notes in Computer Science, pages 1–16. Springer, 2008.

[10] Feikje Hielkema, Chris Mellish, Peter Edwards: **Evaluating an Ontology-Driven WYSIWYM Interface.** INLG 2008.

[11] Davis B, Iqbal AA, Funk A, Tablan V, Bontcheva K, Cunningham H, Handschuh S. **Roundtrip ontology authoring**. In: Proceedings of ISWC 2008. Springer-Verlag; 2008.

Thanks!