Design of a CNL to Involve Domain Experts in Modeling

Sivlie Spreeuwenberg[1], Jeroen van Grondelle[2], Ronald Heller[2], and Gartjan Grijzen[2]

[1] LibRT, Amsterdam, Netherlands
silvie@librt.com
[2] Be Informed, Apeldoorn, Netherlands
{j.vangrondelle, r.heller, g.grijzen}@beinformed.nl

Abstract. Involving domain experts in modeling is important since knowledge needs to be captured in a model and only domain experts can establish whether the models are correct. We have experienced that a natural language based representation of a model helps them to understand the semantics of a model and has advantages over a visual representation. Therefore a controlled natural language (CNL) is designed for our existing semantic reasoning tool based on conceptual graphs (Be Informed). The resulting CNL has a formal logical basis but the goal of the CNL representation is to improve readability for human readers. We report on the challenge to develop a CNL that 1) is easy and intuitively readable for domain experts with no background in formal logics, 2) can be easily generated from the formal representation and 3) can be easily adjusted for other natural languages and cultural preferences. The solution uses patterns to represent the CNL that map to the conceptual graph. The patterns are based on SBVR's RuleSpeak and can be easily adjusted for local differences.

Keywords: Controlled Natural Language, Business Rules, Specifications, Knowledge Representation, CNL design and evaluation, SBVR, RuleSpeak

1 Need for Controlled Natural Languages in Modeling

The adoption of model driven technologies such as Enterprise Decision Management and Business Process Management is growing. As a result, involving business users in modeling is more important than ever. Their ability to capture business knowledge in models correctly is a key factor in the adoption of these technologies. The main challenge in involving business users in knowledge modeling is the fact that most business users are not trained in formal knowledge representation techniques. A formal, concise, visual representation can be quite intimidating to the uninitiated. Consequently they will not be able to verify the accuracy of the model directly.

Be Informed develops a software suite that is used by complex, knowledge intensive organizations to capture their business knowledge and run model driven

services based on these knowledge models. Knowledge representation in Be Informed is based on concept graphs. To add semantics, the concepts, relations and properties are typed, using types from a metamodel associated with the graph. The tool represents the knowledge as a network diagram. A visual syntax maps icons, line styles and colors to metamodel types.

A first version of the textual representation presented in this paper was used to communicate a risk taxonomy to classify shipments of goods to insurance underwriters. Although the sentences produced were very basic and consisted of just the subject and object of a triple with a verb in between encoding for the relation type, the underwriters immediately spotted constructs that appeared odd to them. This resulted in an improved recall rate of modeling errors. This early success has motivated further research at the Dutch Immigration Office [1]. A next version [2] of the text generator was used to validate candidate policy decisions for consistency before they are accepted. In workshops with business representatives and legal advisors, the policy is defined in the tool that also will be used to execute this policy. Both a visual graph oriented representation and the textual representation discussed in this paper were used. It is important to note that the parties involved here were unfamiliar with formal representation techniques and would normally express any policy in unrestricted, natural language. The expectation that the textual representation was preferred over the diagrams was confirmed by the participants. An interesting new observation was that the sentence should be a grammatically correct sentence.

This paper reports on the design and implementation of a CNL that helps Be Informed customers to actively participate in modeling knowledge.

2 Design and Implementation of a CNL for Be Informed

Controlled languages are often classified in one of two categories [3]: those that improve readability for human readers and those that enable reliable automatic semantic analysis of the language. The language that we designed has a formal logical basis. But all too often languages in the second category do not read very naturally. The challenge for Be Informed was to design a language that can be easily generated from a conceptual graph and is natural and understandable for people to read.

Using CNLs to represent ontology's has been done before, for instance in Attempto Controlled Language [4] and CLOnE [5]. They both use natural language generation (NLG) to create a textual representation and natural language processing (NLP) to roundtrip the ontology based on the changed text.

The textual syntax definition proposed in this paper is quite similar to the definition used in CLOnE. Our approach towards editing a model based on a natural language representation does not use NLP and has more in common with Conceptual Authoring [6]. Editing is not performed by manipulating text but by performing editing operations at the concept level, with the text being updated to reflect concept-level changes.

2.1 Pattern Based Generation Approach and SBVR's RuleSpeak

The mechanism we use is based on pattern sentences that map to a concept graph. The formal model remains the single source at all times. The textual representation is just used as a view on the formal model and editing operations by the user in the view are translated into updates to the underlying formal model.

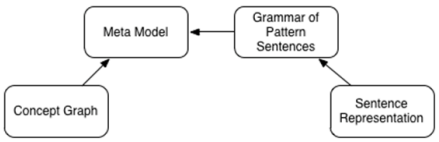

Fig. 1. Pattern based generation approach

The structure of the textual representation of the formal graph is defined by pattern sentences. A pattern sentence consists of static text fragments and subject, object and property placeholders. Fragments and placeholders are grouped into sentence parts, in order to make certain parts of the sentence optional. Cardinality in the (meta) model can be represented using multiple sentences or using enumerated lists of relations of the same type. The sentences are hand-crafted to communicate the semantics of the graph constructs they represent but are re-used for different projects.

The advantage of pattern sentences that map to the formal (meta) model directly is that no NLP or parsing needs to be performed on the textual representation. This provides freedom in choosing or updating the pattern sentences, eventually based on audience-specific preferences, without constraints from NLP techniques. Methods such as RuleSpeak [7] and the OMG standard SBVR [8] have rationalized the use of natural language for the business by introducing syntactic guidelines and best practices. Our latest sentence patterns include RuleSpeak keywords (must, always) and follow the guideline that a rule should be expressed by a grammatically complete, correct and readable sentence. These patterns make sentences easier to read and place a natural and intuitive emphasis on the fact that the sentence introduces an obligation (and is not 'just' a potential statement) for business experts with no background in formal logics. The sentence patterns also provide guidance to direct people into being more formal.

2.2 Implementation and Example

The following sentences illustrate the mapping of pattern sentences to a product model about product purchases and applicable discounts and are taken from a full example in [2]. The sentence parts are placed between quotation marks, the mapping to the metamodel is placed between braces.

This pattern sentence encodes for which discount applies: *"The discount D is always an applicable discount if a customer orders the product P"* {*Discount, requires, Product*} *"with option O."* {*Discount, requires, Option*}.

A rule sentence based on this pattern sentence is: *"The discount early adopters is always an applicable discount for a customer if the customer orders the product basic telephony with option voip."*

Valid rule sentences are created in the editor by choosing relevant pattern sentences and completing the variable parts from a drop-down list. The editor uses the knowledge in the instantiated model and will only present concepts that are defined as a discount in the discount drop-down list.

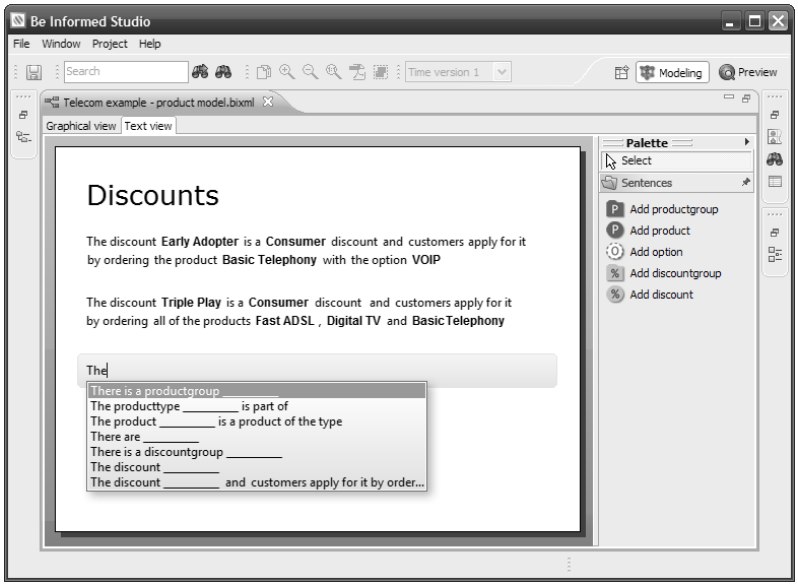

Fig. 2. Task centric and word processor style editing

A consequence of the tight connection between the metamodel and the sentence patterns is that patterns need to explicitly deal with the plural variation of the rule sentence: *"The discount triple play is always an applicable discount for a customer if the customer orders all of the following products:*

- *fast adsl*
- *digital tv*
- *basic telephony."*

Adding knowledge in the algorithm on plural, gender and verb may eliminate this redundancy in the sentence patterns but introduces complexity in the mapping of the sentence to an update on the formal model. It will make the algorithm (natural) language-specific, dependent on the availability of corpora

containing language information on large sets of (often specialized) terminology and results in a more complex user interface for the end-user. These drawbacks have withheld us until now from implementing this strategy.

3 Conclusions and Research Directions

Because no NLP is used in this approach, Be Informed has a lot of freedom in choosing sentence patterns, but has to explicitly deal with grammatical variations. We are interested in hybrid solutions where NLG creates the variations of patterns.

Furthermore, contextualization of the syntax in projects widens the audience (e.g. explanation dialogs, brochures and websites), but increases implementation efforts. To facilitate this trade-off, we need measures that evaluate how well a CNL grammar is suited for an audience. Research in this area and reports on user evaluations (like [9]) are welcome.

References

1. Heller, R., Teeseling, F. van, Gülpers, M.: A Knowledge Infrastructure for the Dutch Immigration Office. 7th Extended Semantic Web Conference, Heraklion. LNCS, vol. 6089, pp. 386-390 (2010)
2. Grondelle, J.C. van, Heller, R., Haandel, E. van, Verburg, T.: Involving Business Users in Formal Modelling using Natural Language Pattern Sentences. Proceedings of EKAW, Lisbon, LNCS (2010)
3. Clark, P., Harrison, P., Murray, W.R., Thompson, J., [ed] Fuchs, N.E.: Naturalness vs. Predictability: A Key Debate in Controlled Languages.Pre-Proceedings of the Workshop on Controlled Natural Language, Marettimo Island. CEUR, vol. 448 (2009)
4. Kaljurand, K., Fuchs, N.E.: Verbalizing OWL in Attempto Controlled English. In Proceedings of Third International Workshop on OWL: Experiences and Directions, Innsbruck, Austria (2007)
5. Funk, A., Tablan, V., Bontcheva, K., Cunningham, H., Davis, B., Handschuh, S.: CLOnE: Controlled Language for Ontology Editing. Proceedings of the Sixth International Semantic Web Conference (ISWC), Busan, Korea, LNCS, vol. 4825, pp. 142-155 (2007)
6. Power, R., Scott, D., Evans, R.: What you see is what you meant: direct knowledge editing with natural language feedback. In Proceedings of the 13th Biennial European Conference on Artificial Intelligence, pp. 675-681, Brighton, UK (1998)
7. Ross, R.G.: RuleSpeak [Online] http://www.rulespeak.com (2009)
8. Object Management Group: Semantics of Business Vocabulary and Rules [Online] http://www.omg.org/spec/SBVR/1.0 (2008)
9. Engelbrecht, P., Hart, G., Dolbear, C. [ed.] Fuchs, N.E.: Talking Rabbit: a User Evaluation of Sentence Production. Pre-Proceedings of the Workshop on Controlled Natural Language, Marettimo Island. CEUR, vol. 448 (2009)