Designing a controlled natural language for the representation of legal norms

Höfler, Stefan ; Bünzli, Alexandra

Designing a Controlled Natural Language for the Representation of Legal Norms

Stefan Hoefler and Alexandra Bünzli

Institute of Computational Linguistics, University of Zurich
{hoefler, buenzli}@cl.uzh.ch

Abstract. Controlled Legal German (CLG) is a controlled natural language being developed for the representation of legal norms contained in Swiss statutes and regulations. This paper discusses the main design requirements CLG faces and the strategies it applies to meet them. CLG aims at providing representations for legal norms that are both formal and can be easily understood and verified by legal experts. It must combine an unambiguous semantics based on FOL and deontic concepts with close syntactic proximity to conventional legal language.

1 Introduction

The development of knowledge-based legal information systems has been at the forefront of research in artificial intelligence and law for the last couple of decades [1, 2]. Knowledge-based legal information systems could support legal experts in constructing and assessing legal arguments or in testing the drafts of new statutes and regulations for consistency and coherence, and they could provide intelligent means for specialised legal information retrieval. One of the main obstacles to the development of knowledge-based legal information systems is the fact that a large part of legal knowledge is captured in natural language texts. For knowledge-based information systems to be employed, these texts must be translated, by hand, into some formal representation [3]. The problem with this requirement is that knowledge engineers do not have sufficient legal expertise to assess the accuracy of these formalisations, while legal experts are usually not familiar with formal representations. Controlled natural language, i.e. the use of a well-defined subset of natural language that has been assigned an unambiguous formal semantics, has been proposed as a means to bridge this gap [4]. Such an approach seems to gain further support from the recent application of controlled natural language to similar domains: contracts [5], business rules [6], clinical guidelines [7].

We are in the process of developing Controlled Legal German (CLG), a logic-based controlled natural language designed for the representation of norms codified in Swiss statutes and regulations. CLG aims at providing representations that are formal *and* can be easily understood and verified by legal experts. In this paper, we will first give a brief overview of the state of development of the project (section 2) and then focus on the specific design requirements that arise

from our target application and the strategies we pursue to meet them in CLG (section 3). The paper concludes with a brief discussion of some problems and limitations and an outlook to future research (section 4).

2 Controlled Legal German

The project is currently in its first phase, in which we specify the syntax and semantics of the controlled language: we define what constructions CLG sentences can consist of and what formal logical representations they map onto. In a second phase, we plan to devise the computational tools needed to automatise the mappings involved: a parser that translates CLG statements into the formal logical representations specified in the definition of the language and a converter that makes the interpretations assumed in CLG explicit by generating respective paraphrases (cf. section 3).

CLG restricts the syntax and semantics of Swiss legal language to prevent instances of ambiguity arising from constructions that either have more than one syntactic analysis (syntactic ambiguity) or whose syntactic analysis can be mapped onto more than one non-equivalent logical structure (semantic ambiguity). Construction rules prohibit the use of certain ambiguous constuctions; interpretation rules assign a default interpretation to others and suggest paraphrases for excluded readings. Like in Attempto Controlled English [8], lexical ambiguity is only controlled in function words; the definition of the meaning of content words is left to the user and to separate ontologies. CLG maps content words onto atomic logical predicates; the often intended vagueness of the legal concepts they stand for is thus fully preserved.

CLG's unambiguous formal semantics is based on first-order predicate logic augmented with defeasible rules ($A \rightsquigarrow B$) and the deontic concepts of obligation ($\mathcal{O}$), permission ($\mathcal{P}$) and prohibition ($\neg\mathcal{P}$). We have chosen a formal underpinning that is expressive and "deep" enough to capture the essential content of individual norms – who must or may do what, how and under what circumstances? – and yet generic enough to be easily converted into other formats of formal representation.

The current version of the language, CLG 1.0, provides the basic syntactic and semantic inventory to express simple norms (obligations, permissions, prohibitions; including norms stating duties and responsibilities). It comprises roughly two dozen construction and interpretation rules that deal with phenomena such as attachment ambiguities (PPs, relative clauses), plural ambiguities (distributive/collective readings), scope ambiguities, lexical ambiguities (function words), referential ambiguities (pronouns, relational nouns) and functional ambiguities due to the relatively free German word order. We will exemplify some of these rules in the next section. The language does not yet include elements of temporal and intensional logic; consequently, it does not yet permit the use of tenses other than present tense and subordinate clauses other than conditional and relative clauses. We plan to add these concepts in CLG versions 2.x and 3.x respectively.

3 Design requirements

The design of CLG has been guided by the two operations it must support: (a) the translation, by hand, of legislative texts into CLG, carried out by knowledge engineers; (b) the verification of the CLG representation by legal experts. The translation of legislative texts into CLG is the easier the closer CLG resembles the language used therein. Moreover, proximity to conventional legal language will make lawyers feel more at ease with CLG transcriptions of statutes and regulations and thus facilitate verification by domain experts. On the other hand, the correctness of CLG transcriptions can only be verified by domain experts if default interpretations are made explicit.

CLG thus needs to accommodate two somewhat converse requirements: (a) proximitiy to conventional legal language and (b) maximal explicitness. It does so by two means. On the one hand, CLG exploits the conventions and frequency distributions of ordinary legal language. Some constructions are ambiguous in full natural language but not in the language of Swiss statutes and regulations, and for some ambiguous constructions, one interpretation is used more frequently in legislative texts than the other. Wherever possible, such conventions and frequency distributions are reflected in CLG's construction and interpretation rules [9]. On the other hand, CLG contains syntactic sugar: for most meanings, multiple constructions are on offer to the user – some of which resemble conventional legal language more closely than others, and some of which are more explicit than others. This fact means that users can first translate a norm into a CLG text that closely resembles the original. This relatively conventional representation can then be transferred deterministically onto a semantically identical but more explicit paraphrase, which can be used to verify the intended interpretation.

As an example, suppose a knowledge engineer needs to formalise the following norm:

(1) Die Bundesversammlung erlässt rechtsetzende Bestimmungen in der Form des Bundesgesetzes [...]. (Art. 163 Abs. 1 Swiss Federal Constitution)

'The Federal Assembly enacts legislative provisions in the form of the federal statute.'

It is not entirely clear whether this is (a) a norm about the responsibilities of the Federal Assembly or (b) a norm about the enactment of legislative provisions by the Federal Assembly. In CLG, the two interpretations can be expressed as shown in (2) and (3) respectively. For each reading, we give a CLG representation that is close to conventional legal language (**C**), an explicit paraphrase equivalent to it (**E**), and the underlying formal semantics[1] (**F**).

(2) **C:** Die Bundesversammlung erlässt rechtsetzende Bestimmungen in der Form des Bundesgesetzes.[2]

[1] For the convenience of the reader, predicate names have been rendered in English.

[2] Note that this CLG representation is positively identical to the original text; the reading it stands for is thus particularly easy to extract.

'The Federal Assembly enacts legislative provisions in the form of the federal statute.'

E: Es ist zwingend, dass die Bundesversammlung eine oder mehrere rechtsetzende Bestimmungen [in der Form des Bundesgesetzes erlässt].
'It is obligatory that the Federal Assembly [enacts in the form of the federal statute] one or several legislative provisions.'

F: $\exists!x : \mathit{federal\_assembly}(x) \wedge$
$\mathcal{O}\ \exists ey : \mathit{legislative\_provision}(y) \wedge \mathit{enacts}(e, x, y) \wedge$
$\mathit{in\_form\_of\_federal\_statute}(e)$[3]

(3) **C:** Der Erlass von rechtsetzenden Bestimmungen durch die Bundesversammlung erfolgt in der Form des Bundesgesetzes.
'The enactment of legislative provisions by the Federal Assembly is effected in the form of the federal statute.'

E: Es ist zwingend, dass jeder Erlass einer rechtsetzenden Bestimmung durch die Bundesversammlung in der Form des Bundesgesetzes erfolgt.
'It is obligatory that every enactment of a legislative provision by the Federal Assembly is effected in the form of the federal statute.'

F: $\exists!x : \mathit{federal\_assembly}(x) \wedge$
$\mathcal{O}\ \forall ey : \mathit{enacts}(e, x, y) \wedge \mathit{legislative\_provision}(y) \rightarrow$
$\mathit{in\_form\_of\_federal\_statute}(e)$

The paraphrases make the following CLG default interpretations underlying the conventional CLG representations explicit: (a) norms without modal verb are considered obligations (by preceding the sentence with the phrase *es ist zwingend, dass*), (b) PPs attach to the verb (by inserting brackets), (c) indefinite NPs in non-vorfeld position are interpreted as existentially quantified *rechtsetzende Bestimmungen* (by using determiners such as *eine oder mehrere*), and (d) nominalised verbs in vorfeld position (here *der Erlass*) are interpreted as universally quantified (by adding the determiner *jeder*).

4 Discussion

The fact that legislative language is highly conventionalised facilitates the task of designing a resembling controlled natural language to some degree [9]: ambiguous constructions often already have a default interpretation in full legislative language (due to the pragmatics of the domain), and these interpretations can be turned into interpretation rules in the controlled natural language. What makes our work more difficult, however, is the lack of thorough linguistic studies of how specific potentially ambiguous constructions (e.g. indefinite NPs, plurals, etc.) are used in legislative language.

The provision of explicit paraphrases often poses further problems: it is simply not possible to find explicit paraphrases for all constructions to which CLG applies an interpretation rule. We have tried to partially overcome this problem

[3] Adverbial phrases are not yet analysed in CLG 1.0 but treated as atomic conditions.

by admitting non-linguistic elements in these paraphrases. An example are the brackets we use in (2**E**) to indicate the attachment of constituents.

The foremost limitation of our project lies in the fact that the field of artificial intelligence and law has not yet arrived at developing a logical formalism capable of adequately representing the full content of statutory law. For the moment, we have therefore chosen a formal underpinning for CLG that is expressive and "deep" enough to capture what we deem the essential content of individual norms – who must or may do what, how and under what circumstances? – and yet generic enough to be easily converted into other formats of formal representation. We believe that with such a generic representation, chances are best that a future integration into artificial intelligence and law systems that model legal reasoning will be possible. Such systems will, of course, also have to take into account information that is not contained in the actual norms that we want to represent in CLG. Examples are the precedence of more specific norms over more general norms or the precedence of constitutional provisions over provisions stated in statutes or regulations.

Finally, the notorious underspecificity of normative stentences continues to pose a problem for their translation into a controlled natural language: many relations within sentences and between sentences (bridging references, ellipses, rule-exception relations) are not made explicit in normative texts but are left to be inferred from the context. In a controlled natural language, such relations have to be stated explicitly. Providing formalisms for doing so will be one of the largest task to be tackled by our future research.

References

1. Rissland, E.L., Ashley, K.D., Loui, R.P.: AI and law: A fruitful synergy. Artificial Intelligence **150**(1–2) (2003) 1–15
2. Lodder, A.R., Oskamp, A.: Information Technology & Lawyers: Advanced technology in the legal domain, from challenges to daily routine. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006)
3. McCarty, L.T.: Deep semantic interpretation of legal texts. In: Proceedings of the 11th Int. Conference on AI and Law, New York, ACM Press (2007) 217–224
4. Hoey, M., Walter, C.: Natural language interfaces. In Walter, C., ed.: Computer Power and Legal Language, New York, Quorum (1988) 135–142
5. Pace, G., Rosner, M.: A controlled language for the specification of contracts. In Fuchs, N., ed.: Controlled Natural Language, Berlin, Springer (2010)
6. Spreeuwenberg, S., Anderson Healy, K.: SBVR's approach to controlled natural language. In Fuchs, N., ed.: Controlled Natural Language, Berlin, Springer (2010)
7. Shiffman, R.N., Michel, G., Krauthammer, M., Fuchs, N.E., Kaljurand, K., Kuhn, T.: Writing clinical practice guidelines in controlled natural language. In Fuchs, N.E., ed.: Controlled Natural Language, Springer (2010)
8. Fuchs, N., Kaljurand, K., Kuhn, T.: Attempto Controlled English for knowledge representation. In Baroglio, C., Bonatti, P., Maluszynski, J., Marchiori, M., Polleres, A., Schaffert, S., eds.: Reasoning Web, Berlin, Springer (2008) 104–124
9. Hoefler, S., Bünzli, A.: Controlling the language of statutes and regulations for semantic processing. In: Proceedings of the LREC 2010 Workshop on Semantic Processing of Legal Texts (SPLeT 2010), Valletta, Malta (2010) 8–15