

Controlling the language of statutes and regulations for semantic processing

Stefan Hoefler and Alexandra Bünzli

University of Zurich, Institute of Computational Linguistics
Binzmühlestrasse 14, 8050 Zurich, Switzerland
hoefler@cl.uzh.ch, buenzli@cl.uzh.ch

Abstract

Controlled Legal German (CLG) is a subset of legal German specifically designed to facilitate the semantic processing of Swiss statutes and regulations. In this paper, we describe the strategies CLG employs to reduce ambiguity and underspecification in such texts, and the methods it uses to maintain proximity to conventional legal language. The presented discussion suggests that, if existing synergies are properly exploited, the concept of controlled natural language can be of benefit to the semantic processing of legal texts as well as to legislative drafting.

1. Introduction

The last two decades have brought substantial progress in the development of formal logical representations of legal knowledge and of methods to perform automated legal reasoning with these representations (Rissland et al., 2003). However, as McCarty (2007, p. 217) observes,

[o]ne of the main obstacles to progress in the field of artificial intelligence and law is the natural language barrier. Since the raw materials of the law are embodied in natural language – cases, statutes, regulations, etc. – the designer of a knowledge-based legal information system today must translate them, by hand, into a formal language, just to get started.

Since a manual translation of legal texts into formal logical representations is both time-consuming and error-prone, the employment of natural language processing techniques seems to be the only viable option to bridge the gap between legal texts and knowledge-based legal information systems. While state-of-the-art methods of natural language processing have come to deliver fairly decent results (McCarty, 2007), they continue to struggle with the notoriously difficult resolution of natural language ambiguity and underspecification.

The Collegis project (Controlled Language for Legal Information Systems) addresses this problem from the perspective of legislative drafting. We develop Controlled Legal German (CLG), a restricted version of Swiss legal German specifically designed to facilitate the semantic processing of statutes and regulations.

Controlled languages restrict the vocabulary, syntax and/or semantics of a natural language in order to reduce its ambiguity and complexity. While early versions of controlled languages were mainly devised to improve the readability and translatability of texts, recently, the method has been used to define subsets of natural languages that can be unambiguously translated into formal logic (Pool, 2006; Fuchs et al., 2008). Controlled languages have been developed for the domains of technical documentation and requirements engineering and for general-purpose knowledge representation. There have also been first attempts to apply the method to defining business rules (Spreeuwenberg

and Anderson Healy, 2009) and writing contracts (Pace and Rosner, 2009).

In this paper, we build on a proposal by Hoey and Walter (1988) and introduce legislative drafting as another promising area of application. Legislative drafting, by definition, already exerts a certain degree of control on legal language, thereby pursuing aims similar to those of controlled languages: the reduction of ambiguity and sufficient specification of rules. While there have been studies on improving the understandability of legal language (Wydick, 2005; Neumann, 2009), no controlled legal language has as yet been developed for the purpose of facilitating automated semantic processing.

The remainder of the paper is organized as follows. We first give an overview of the rationale behind CLG and the methods it applies. After detailing the aims of CLG (section 2), we introduce the methods it uses (section 3) and illustrate with a specific example how these methods are applied (section 4). Afterwards, we demonstrate how CLG exploits conventions that already exist in Swiss legal language (section 5) and discuss approaches to controlling underspecification in statutes and regulations (section 6). After describing the current state of development of CLG (section 7), we conclude with the presentation of a brief proposal for the evaluation of controlled legal languages (section 8) and with a discussion of the potential and limitations of the approach for both semantic processing of legal texts and legislative drafting (section 9).

2. Aim

The goal we pursue with the development of CLG is to provide a language for Swiss statutes and regulations whose semantics can be understood by humans and processed by computers. To allow for an automatic translation of such texts into formal logical representations which can e.g. be fed to some automated inference system, CLG aims at reducing natural language ambiguity. However, while CLG eliminates lexical ambiguity in function words and law-specific expressions, it leaves the interpretation of content words to the terminology databases and ontologies of its users. CLG does therefore not infringe on the often intended vagueness and open-textured nature of the concepts represented by content words (Gardner, 1987). Like other

controlled languages that aim at providing an interface to some sort of formal logic, such as ACE (Fuchs et al., 2008) or PENG (Schwitter and Tilbrook, 2006), CLG is mainly concerned with the reduction of syntactic and semantic ambiguity.

Syntactic ambiguity occurs in situations where a sentence can be assigned more than one syntactic structure. Typical examples are so-called attachment ambiguities: in sentence (1), the prepositional phrase *im Bereich der Logistik* ('in the sector of logistics') could theoretically be attached to *deckt* ('supplies'), to *Bedarf* ('need'), to *Güter und Dienstleistungen* ('goods and services'), or only to *Dienstleistungen* ('services').

- (1) Das Bundesgericht deckt seinen Bedarf an Gütern und Dienstleistungen im Bereich der Logistik selbständig. (Art. 25a Abs. 2 BGG¹)
'The Federal Supreme Court supplies its need for goods and services in the sector of logistics autonomously.'

Semantic ambiguity, on the other hand, occurs if a sentence has only one syntactic structure but can be assigned two or more non-equivalent logical representations. A typical example is scope ambiguity. Without contextual knowledge, it cannot be determined whether sentence (2) requires that each representative (or rather the representatives of each party) show a separate letter of attorney, or whether it means that they should provide one letter of attorney together.

- (2) Die Parteivertreter und -vertreterinnen haben sich durch eine Vollmacht auszuweisen. (Art. 40 Abs. 2 BGG)
'The party representatives have to identify themselves with a letter of attorney.'

Besides reducing ambiguity, CLG aims at preventing types of underspecification that warrant unintended inferences. This goal is described in more detail in section 6.

3. Methods

Generally, controlled natural languages use the following methods to reduce natural language ambiguity and complexity:

- *Construction rules*
Construction rules restrict the number of words and constructions that can be used, thus prohibiting the use of specific ambiguous words and constructions.
- *Interpretation rules*
Interpretation rules assign default interpretations to the remaining ambiguous words and constructions.
- *Paraphrases*
Paraphrases suggest alternative ways of expressing the respective other meaning of an originally ambiguous word or construction.

¹Bundesgerichtsgesetz (Federal Supreme Court Act), SR 173.100

One of the main problems of the method of controlled natural language is the fact that there is a trade-off between the level of control a language exhibits (and thus its processability) and its expressiveness, naturalness and user-friendliness. Most existing controlled natural languages, especially those aspiring to provide an interface to some kind of formal logic, consequently have only very limited expressiveness, and several of them include constructions which border on naturalness at best (Pool, 2006). CLG differs from these languages as it needs to be expressive enough to render the contents of statutes and regulations, and natural enough to be understood and accepted by non-expert human readers. We employ three methods to maximize CLG's proximity to ordinary legal language:

- *Syntactic sugar*
As the naturalness of specific control mechanisms may vary from context to context, CLG usually provides more than one way of controlling an individual phenomenon.
- *Variable-depth control*
For certain phenomena, CLG provides multiple levels of control, which can be switched on or off by the user, depending on the requirements of the target application.
If further specificity is not required, certain ambiguous constructions are only assigned underspecified logical representations. The treatment of such constructions is then left to the tools that process the logical representations.²
- *Interactive control*
Some phenomena are not controlled statically but resolved dynamically by providing an authoring tool that asks the user to specify the intended meaning upon each occurrence of the respective ambiguous construction (Macias and Pulman, 1995).
To guarantee transparency, the choices made by the user are recorded in a so-called disambiguation protocol, which is to be stored together with the text.

In the next section, we illustrate with a specific example how these methods are applied.

4. Applying the methods

In Swiss statutes and regulations, indefinite noun phrases in subject position usually indicate what the respective norm is about, i.e. they introduce the "subject matter" of the norm (Caussignac et al., 2000). However, the indefinite article (*ein/eine/ein* 'a(n)' in singular; \emptyset in plural) is ambiguous at this position: it can have an existential interpretation, as in example (3), or a generic interpretation, as in example (4), which can be represented as universal quantification (Gamut, 1991; Cohen, 2001).³

²Attempto Controlled English (Fuchs et al., 2008) uses this method for plural ambiguities and copula; Computer Processable Language (Clark et al., 2005) employs it for PP-attachment ambiguities.

³In statutes and regulations, the generic interpretation of the indefinite article does not express the prototypical features of a

- (3) Ein Mitglied der Universitätsleitung führt den Vorsitz.
(§ 67 Abs. 2 UniO UZH⁴)

‘A member of the Executive Board of the University acts as chair.’

$\exists x : member(x) \wedge \dots$

- (4) Ein Titel [...] kann von der Erweiterten Universitätsleitung auf Antrag der Fakultät entzogen werden, wenn die Inhaberin oder der Inhaber die Interessen der Universität ernsthaft verletzt.
(§ 8 Abs. 7 UniO UZH)

‘A title [...] can be revoked by the Extended Executive Board of the University at the request of the faculty if the holder seriously violates the interests of the university.’

$\forall x : title(x) \rightarrow \dots$

One way to control this ambiguity is to define a construction rule that prohibits the use of the indefinite article *ein* altogether and to offer paraphrases for its two interpretations in order to maintain the expressiveness of the language. An existentially quantified subject matter could be introduced by *mindestens ein* (‘at least one’) or *genau ein* (‘exactly one’); for a universally quantified subject matter, one could use the determiner *jeder* (‘every’). The problem with this solution is that it represents a significant deviation from conventional legal language, where the use of *ein* is very common while the use of *mindestens/genau ein* and *jeder* is rare and more marked. Adopting this construction rule would thus substantially decrease the naturalness of CLG. The solution at the other end of the scale is to resolve the ambiguity caused by the use of *ein* interactively, i.e. to devise an authoring system that asks the user to indicate for every occurrence of a subject matter introduced by *ein* whether existential or universal quantification is intended. However, while interactive control is a viable option for relatively rare phenomena, it is clearly not user-friendly for phenomena that occur as frequently as the indefinite article. This solution too must be rejected.

The only remaining method of control is the definition of an interpretation rule that identifies one of the two readings of *ein* as the default interpretation. In statutes and regulations, indefinite plural noun phrases in subject position generally exhibit a generic reading and are thus to be represented as universally quantified. Sentence (5) provides an example.

- (5) Dienstleistungen sind in der Regel mindestens kostendeckend in Rechnung zu stellen.
(§ 3 Abs. 3 UniO UZH)

‘Services usually have to be charged so that at least the costs covered.’

$\forall x : service(x) \rightarrow \dots$

kind but states a rule that applies to every instance of that kind.
– This observation can be conceived as a CLG interpretation rule defining that the generic reading of the indefinite article is interpreted as universal quantification.

⁴Universitätsordnung der Universität Zürich (University Regulation of the University of Zurich), SR 415.111

As CLG aims at staying close to conventional legal language, it would make little sense to define the existential reading as the default interpretation of indefinite plural noun phrases. For indefinite singular noun phrases, neither interpretation can be considered conventional. To keep the number of rules that users of CLG need to master low, we apply one and the same interpretation rule to both the singular and the plural version of the indefinite article: indefinite noun phrases are interpreted as universally quantified in subject position (and as existentially quantified elsewhere; see section 7).

The definition of such an interpretation rule entails that example (3) needs to be re-phrased to obtain existential quantification. Two options are available. The first is to make the existential quantification explicit by using determiners such as *mindestens ein* (‘at least one’) or *genau ein* (‘exactly one’). In the present example, however, these determiners do not sound particularly natural and potentially confuse the reader as they seem to be marked pragmatically:

- (6) Genau ein Mitglied der Universitätsleitung führt den Vorsitz.

‘Exactly one member of the Executive Board of the University acts as chair.’

Alternatively, the noun phrase *ein Mitglied der Universitätsleitung* can be moved away from the subject position. This effect can be achieved by using a passive construction such as (7). For the present example, this second solution provides a sentence that both feels natural and is interpreted in the intended way in CLG.

- (7) Die Forschungskommission wird von einem Mitglied der Universitätsleitung präsiert.

‘The research committee is chaired by a member of the Executive Board of the University.’

Sentence (7) is preferable to (3) not just from the perspective of semantic processing but also from the perspective of legislative drafting. First, the subject of a norm should usually indicate what this norm is about. The present norm is not about some member of the Executive Board of the University but about the research committee. Second, the rephrased version indicates explicitly what the chair is *of* (namely the research committee); in the original version, this information has to be inferred from the context. We come back to such phenomena of underspecification in section 6.

Depending on their target application, some users of CLG may not want to commit to the aforementioned interpretation rules. Answer extraction, for instance, can cope without the explicit specification of quantification (Mollá, 2001). As CLG pursues a policy of variable-depth control, it therefore also provides the option of leaving the quantification of indefinite noun phrases underspecified. In that case, the aforementioned interpretation rules do not apply.

5. Exploiting domain-specific conventions

Since its aims are similar to those of controlled natural language, conventional legal language itself provides mechanisms to control certain types of ambiguity. Whenever possible, CLG exploits these already existing mechanisms.

CLG makes, for instance, use of the fact that some words and constructions that are ambiguous in full natural language have acquired a default interpretation in legal language. In ordinary German, the adverb *grundsätzlich*, modifying an obligation or permission, can have two directly opposed interpretations: if interpreted in the sense of ‘strictly’ or ‘categorically’, it denotes that the respective rule does not allow for exceptions; if interpreted as ‘generally’ or ‘in principle’, it indicates that the rule allows for exceptions, which is particularly relevant in the context of defeasible reasoning. By convention, *grundsätzlich* is always used in the latter sense in Swiss legal German. CLG therefore devises an interpretation rule defining that *grundsätzlich* is always interpreted as indicating the admissibility of exceptions:

- (8) Die Veröffentlichung der Entscheide hat grundsätzlich in anonymisierter Form zu erfolgen. (Art. 27 Abs. 2 BGG)
 ‘In principle, the decisions must be published in anonymized form.’

Note that unlike ordinary adverbs, *grundsätzlich* does not modify the verb but the obligation as a whole. CLG defines a number of words and fixed expressions that are not interpreted like other items of the same grammatical category but obtain domain-specific interpretations. Table 1 lists the most common of them.

Another example of a phenomenon for which CLG exploits existing domain-specific methods of control is attachment ambiguity in complex coordination structures. Sentences like (9) are difficult to parse not only for computers but also for humans. It is thus in the best interest of both NLP and legislative drafting to control the attachment ambiguities they contain.

- (9) In Fünferbesetzung entscheiden sie ferner über Beschwerden gegen referendumspflichtige kantonale Erlasse und gegen kantonale Entscheide über die Zulässigkeit einer Initiative oder das Erfordernis eines Referendums. (Art. 20 Abs. 3 BGG)
 ‘In a composition of five, they furthermore decide on appeals against cantonal decrees subject to referendum and against cantonal decisions on the admissibility of an initiative or the necessity of a referendum.’

CLG defines an interpretation rule stating that constituents are always attached to the closest possible candidate. While this rule can be easily applied to relatively simple sentences, it is clearly not user-friendly enough, both in terms of writability and readability, for complex coordination structures such as (9). To disambiguate such structures, CLG includes a means provided by conventional legal language: ellipses are removed by repeating all elements in each conjunct (in this case, the phrase *kantonale Entscheide* ‘cantonal decisions’ is repeated) and the conjuncts are listed in enumerations introduced by letters, as shown in (10).

- (10) In Fünferbesetzung entscheiden sie ferner über Beschwerden gegen:

- a. referendumspflichtige kantonale Erlasse;
- b. kantonale Entscheide über die Zulässigkeit einer Initiative;
- c. kantonale Entscheide über das Erfordernis eines Referendums.

‘In a composition of five, they furthermore decide on appeals against:

- a. cantonal decrees subject to referendum;
- b. cantonal decisions on the admissibility of an initiative;
- c. cantonal decisions on the necessity of a referendum.’

6. Controlling underspecification

Besides ambiguity, underspecification is the main issue that a controlled legal language needs to address. We can distinguish two types of underspecification in statutes and regulations.

The first type occurs where legislators deliberately refrain from specifying certain details. Sentence (11) may serve as an example.

- (11) Die Bundesversammlung wählt die Richter und Richterinnen. (Art. 5 Abs. 1 BGG)
 ‘The Federal Assembly elects the judges.’

In general, plural noun phrases can have a distributive reading (each judge is elected individually) and a collective reading (the judges are elected as a body).⁵ As the distributive interpretation is far more frequent in statutes and regulations, CLG defines it as the default interpretation. To express the collective reading, a singular term has to be used (e.g. *das Gericht* ‘the court’). This strategy can also be frequently found in existing legal texts. However, even with such an interpretation rule being applied, sentence (11) remains indeterminate: it does not specify the exact conditions under which an individual judge is considered elected. Even if the Federal Assembly elected the judges as a body, each judge might be considered elected individually by this act. The legislator deliberately leaves the conditions under which a judge needs to be elected undetermined here; CLG reflects this fact despite the application of an interpretation rule.

The second type of underspecification poses a much more substantial problem to the semantic processing of statutes and regulations. It occurs in passages that warrant unintended inferences if they are not further specified and that are therefore potentially harmful to correct automated reasoning. Sentence (12) is an example.

- (12) Bei der Geburt eines Kindes hat der Angestellte Anspruch auf eine einmalige Zulage von 530 Franken. (Art. 55 Abs. 1 AngO ETH-Bereich⁶)
 ‘Upon the birth of a child, the employee is entitled to a one-time allowance of 530 francs.’

⁵The treatment of plural ambiguities in controlled language is thoroughly discussed in Schwertel (2000).

⁶Angestelltenordnung ETH-Bereich (Employee Regulation ETH-Domain), SR 172.221.106.2

Word / Expression	Translation	Function
muss/müssen, hat/haben zu*	must, have to	marks a rule as an obligation
darf/dürfen, kann/können*	may, can	marks a rule as a permission
grundsätzlich, in der Regel*	in principle	indicates that a rule allows for exceptions
gemäss, im Rahmen (von)	according to, within the scope (of)	indicates the applicability of another rule
(gilt) sinngemäss	(applies) analogously	indicates that a rule is applicable in adapted form
insbesondere, namentlich*	in particular	indicates that a specific case is made explicit
bei (+NP)	upon (+NP), ~if	condition in the form of a PP; cf. example (12)

Table 1: Examples of function words and fixed expressions with conventionalized domain-specific interpretations. (Expressions marked with an asterisk are contained in CLG 1.0; cf. section 7.)

The problem sentence (12) poses for semantic processing is that the condition (*at the birth of a child*) apparently does not share any discourse referent with the consequence (*the employee is entitled to a one-time allowance of 530 francs*). The sentence does not specify explicitly that the employee does not receive an allowance on the occasion of the birth of just *any* child but only if he or she is the parent of that child. Human readers will easily infer this missing bit of information from the context and thus reduce the number of warranted inferences. An automated reasoner, on the other hand, may in the worst case combine the logical representation of (12) with the fact that approximately 216,000 children are born every day, and deduce that an employee is to receive total allowances of 114,480,000 francs per day. To avoid this problem, a controlled legal language may prescribe that the condition of a conditional norm always has to share a discourse referent with its consequence. This requirement can be fulfilled, for instance, by augmenting the condition with a relative clause:

- (13) Bei der Geburt eines Kindes, *gegenüber dem er elterliche Pflichten hat*, hat der Angestellte Anspruch auf eine einmalige Zulage von 530 Franken.
‘Upon the birth of a child *toward whom he or she has parental duties*, the employee is entitled to a one-time allowance of 530 francs.’

The same effect is achieved if another condition is added at the end of the sentence:

- (14) Bei der Geburt eines Kindes hat der Angestellte Anspruch auf eine einmalige Zulage von 530 Franken, *sofern er gegenüber dem Kind elterliche Pflichten hat*.
‘Upon the birth of a child, the employee is entitled to a one-time allowance of 530 francs, *provided that he or she has parental duties toward the child*.’

Note that the application of this rule is not only beneficial to semantic processing but also to legislative drafting. Had they been forced to provide the additional specification, legislators would have become aware of an overlooked regulatory loophole, namely that biological parents who are not liable for support should not be entitled to an allowance while foster parents should.

An alternative solution to controlling the phenomenon becomes available if one recognizes that the noun *Kind* (‘child’) is in fact ambiguous: it can denote a young human,

or it can denote someone’s direct offspring. In the latter sense, *Kind* is a relational noun, whose logical representation takes not one but two arguments: *child_of(x, y)*. The noun *Kind* is thus implicitly anaphoric, referring to some other entity in the text.⁷ The problem is then to constrain the field of potential antecedents so that the implicit reference is unambiguous. The guidelines for legislative drafting provided by the Swiss Confederation (BJ, 2007) and by the Canton of Zurich (ZH, 2005) constrain the use of pronouns – another type of anaphoric references – in statutes and regulations: pronouns may only refer to entities within the same article and they may only refer to either the subject of the main clause or to the subject of the immediately preceding sentence. The same rule can now be applied to the implicit anaphoric references created by relational nouns: their use may be constrained to referring to either the subject of the main clause (as is the case in our example) or, if they are part of that subject themselves, to the subject of the immediately preceding sentence – provided that that sentence is in the same article as the sentence containing the relational noun. It needs to be said, however, that this rule is not yet implemented in CLG 1.0, the version of the language representing the current state of development, which we will briefly describe in the next section.

7. State of development

The state of development of Controlled Legal German is reflected in version 1.0 of the language, which is documented in Hoefler and Bünzli (2010). CLG 1.0 provides the basic syntactic and semantic inventory to express simple norms (obligations, permissions, prohibitions; including norms stating duties and responsibilities) as well as legal definitions. Example (15) provides a typical CLG 1.0 sentence and the logical representation it maps onto.

- (15) Radfahrer müssen mindestens zwei rote Rückstrahler tragen, sofern sie keine Ausnahmegewilligung haben.
‘Cyclists must wear at least two red reflectors, unless they have (if they do not have) a certificate of exception.’

$$\begin{aligned}
\mathcal{O} \quad \forall x : & \quad [\text{radfahrer}(x) \wedge \\
& \quad \neg \exists y : [\text{ausnahmegewilligung}(y) \wedge \\
& \quad \exists e : \text{has}(e, x, y)]] \\
\rightarrow & \quad \exists^{\geq 2} z : [\text{roter_rueckstrahler}(z) \wedge \\
& \quad \exists f : \text{traegt}(f, x, z)]
\end{aligned}$$

⁷We thank an anonymous reviewer for pointing this out.

Sentence (15) illustrates the following characteristics of the formal semantics underlying version 1.0 of Controlled Legal German. CLG 1.0 can be unambiguously mapped onto predicate logic representations that are augmented with deontic operators for obligation (\mathcal{O}), permission (\mathcal{P}) and prohibitions ($\neg\mathcal{P}$). Since content words are translated into predicate symbols, the potential open-texturedness of the concepts they represent is preserved. The events and states represented by verbs are reified (and quantified). Adjectives used in an attributive manner and the nouns they modify are, for now, contracted into a single logical predicate. CLG 1.0 includes means to express existential and universal quantification, as well as counting quantifiers. It does, however, not include elements of temporal and intensional logics. These concepts are planned to be added to the standard in CLG versions 2.x and 3.x respectively. The following list provides an overview of the range of syntactic constructions that are available in CLG 1.0:

1. Only present tense is permitted.
2. Sentences have canonical word order (the subject preceding objects and adverbials).
3. Both active and passive voice are permitted.
4. Nouns can currently be modified by (a) adjectives, (b) participle constructions, (c) relative clauses, but not by prepositional phrases (with the exception of the prepositional phrase denoting the agent of a nominalized verb).
5. Verbs can be modified by (a) adverbs, (b) prepositional phrases.
6. Main clauses may contain a modal verb; main clauses without modal verb are assumed to be obligations.
7. Nouns, verbs and adjectives can be coordinated; coordinations may be put in the form of enumerations (cf. section 5).
8. Attributes in genitive case are only permitted to express the direct object of nominalized verbs or the complements of relational nouns.
9. Conditional clauses and relative clauses are the only permissible subordinate clauses.
10. Complement clauses and adverbial clauses are not permitted (with the exception of conditional clauses).
11. There are special formulaic expressions to list duties and responsibilities, such as e.g. *X hat die folgenden Aufgaben und Kompetenzen* ('X has the following duties and responsibilities').

The semantics of CLG 1.0 sentences is controlled by the following seven interpretation rules:

1. Modal verbs have wide scope over the rest of the sentence.
2. Subjects have wide scope over Objects and Adverbials.

3. Pronouns refer to the subject of the sentence or, if they are part of the subject, to the subject of the immediately preceding sentence.
4. Indefinite noun phrases are interpreted generically if they are the subject of a sentence and existentially elsewhere.
5. Plurals are interpreted distributively. If a collective reading is intended, a singular term has to be used.
6. Definite noun phrases presuppose existence and uniqueness and are interpreted referentially. Definite plurals are interpreted distributively and universally.
7. Attachment ambiguity is resolved by attaching the constituent in question to the closest potential antecedent; if that antecedent is a conjunct, the constituent is attached to the whole coordination.

For a detailed account of CLG 1.0, we refer to Hoefler and Bünzli (2010). Evidently, CLG 1.0 is not yet expressive enough to be used in legislative drafting. It can, however, be employed to model simple norms in a way that provides a formal specification and is yet understandable by non-expert human readers.

8. Proposal for evaluation

As the development of CLG is still work in progress, a thorough evaluation of the controlled natural language can not yet be provided. In any case, before such an evaluation can be undertaken, one needs to define how a controlled language that aims at facilitating the semantic processing of statutes and regulations is to be assessed in the first place. We propose that such a controlled language has to be evaluated for the following criteria:

- *Expressiveness*
An ideal controlled legal language should be able to express all propositions that conventional legal language can express. The expressiveness of a controlled legal language can be assessed by determining what percentage of the content of a chosen statute or regulation (e.g. how many of the individual norms contained in that text) it can express.
- *Proximity to conventional legal language*
An ideal controlled legal language should be indistinguishable from conventional legal language in terms of style. As a first approximation, the degree to which a controlled legal language covers conventional legal language can be evaluated by assessing how many articles of a chosen statute or regulation need to be altered if that text is to be translated into the controlled language. If only few passages have to be altered, the respective controlled language can be considered stylistically close to conventional legal language. However, the need for rephrasing does not necessarily imply that the resulting text deviates from the conventions of legal language. It may still be perfectly acceptable. The stylistic acceptability of substantially altered texts therefore requires additional assessment by human legal editors.

As we have already pointed out above, the degree to which a controlled language covers conventional legal language can be maximized by the use of syntactic sugar, the employment of control mechanisms that reflect the frequency distributions in the reference texts, and the provision of authoring systems that allow for certain phenomena to be disambiguated interactively.

Both criteria, expressiveness as well as proximity to conventional legal language, can only be assessed in a controlled legal language if a corpus of semantically analyzed reference texts is available. We are currently building such a corpus, starting with the Federal Supreme Court Act and the Regulation of the University of Zurich, from which we quoted in this paper.

9. Discussion and conclusion

The conditions encountered in legislative drafting seem ideal for an application of controlled natural language. Statutes and regulations are written in a highly conventionalized language that contains restrictions aimed at preventing ambiguity. Due to this shared goal, the properties of legal language are not unlike those of typical controlled natural languages. Controlled Legal German uses these synergies to facilitate the drafting of statutes and regulations that can be automatically translated into formal logical representations. It thus attempts to bridge the gap between legal texts, written in natural language, and knowledge-based legal information systems, operating with formal logical representations.

In this paper, we have presented the general rationale behind CLG and introduced the methods it applies to prevent ambiguities and underspecification. We have shown that CLG consists of a set of recommendations in the form of construction and interpretation rules of variable depth, accompanied by suggested paraphrases and options for interactive ambiguity resolution. These recommendations explain how statutes and regulations can be formulated in a way that enables automatic semantic processing. In parallel to defining such rules, we are working on combining them into a comprehensive formal description of a subset of Swiss legal German that can be translated deterministically into formal logical representations.

At this point, some remarks on the limitations of the approach of applying controlled natural language to legal texts are in place. A first limitation of the approach pertains to the availability of adequate logical representations for the content expressed linguistically in norms. Not all linguistic phenomena can easily be represented in formal logic: temporal relations or intensional contexts, for instance, already require a rather complex machinery of operators and axioms. But even apparently simple linguistic constructions such as attributive genitives or opaque adjectives do not have straightforward logical representations. There will always be some phenomena that have to be treated as “black boxes” or modeled in a grossly simplified manner. Any logical representation derived from a norm written in controlled natural language will thus only capture the content of that norm to a certain degree of granularity.

A second limitation of the approach refers to the fact that controlled natural languages such as CLG, ACE or PENG

may be able to reduce (or, in the ideal case, eliminate) certain types of ambiguity but cannot (and do not aim to) remove vagueness. The predicates of any logical representation will still stand for concepts whose definition may be vague or open-textured. On the one hand, this fact reflects a reality of legal language, where vagueness and indeterminacy is often positively intended (Nussbaumer, 2005). On the other hand, it means that being able to derive a logical representation from a legal text written in a controlled language does not entail that one will automatically be able to perform meaningful legal reasoning over such a representation. While certain inferences can be drawn purely on the basis of the logical representations of a statute or regulation by treating the logical predicates and the potentially vague concepts they stand for as black boxes, deeper automated reasoning will in addition at least require extensive ontologies modeling world knowledge.

A third limitation pertains to the controlled natural language itself. It is to be expected that extensive control of ambiguity will lead to a certain reduction of the expressiveness of legal language. The future development of CLG will have to show whether this reduction can be kept at a level at which it does not seriously impede the usability of CLG for legislative drafting. While experience shows that many types of ambiguity can be controlled by the methods described, underspecification will continue to pose a problem. It will most likely not be possible for a controlled language to prevent the vast number of situations in which a human writer may underspecify some of the information required for accurate reasoning.

Finally, the success of a controlled legal language will depend on its acceptance by professional legal editors: CLG must be easy to learn and close to conventional legal language both in terms of the propositions it can express and the stylistic means it provides. It is too early to speculate if it will be possible to develop a controlled version of legal German that is accepted by its potential users. Using CLG for didactic purposes and for the conceptualization of norms rather than for actual legislative drafting may be a fallback strategy. However, there are three factors that at least have the potential to exert a positive influence on the acceptability of employing controlled natural language in legislative drafting. First, professional legal editors are domain-specialists that are used to (and well capable of) following linguistic guidelines. The application of such rules may be additionally supported by specifically designed authoring tools (Schwitter et al., 2003). Second, there is some chance that we will be able to show that the employment of a controlled legal language can also be beneficial to legislative drafting itself. In this paper, we have briefly demonstrated how CLG can help legal editors detect regulatory loopholes they would otherwise have overlooked. Eventually, however, the acceptance of controlling legal language for semantic processing will be served best if it grants access to AI & Law applications of evident practical use beyond legislative drafting.

Acknowledgements. The authors wish to thank the Central Language Services of the Swiss Federal Chancellery for

their advice on the conventions of legal language and the process of legislative drafting, four anonymous reviewers for their valuable comments and suggestions, and Michael Hess for his scientific and institutional support of the Col-legis project.

10. References

- BJ (Bundesamt für Justiz), Bern, 2007. *Gesetzgebungsleit-faden: Leitfaden für die Ausarbeitung von Erlassen des Bundes*, 3. edition.
- G. Caussignac, C. Eberhard, P. Häusler, D. Kettiger, D. Pulitano, and R. Schneider, 2000. *Rechtsetzungs-richtlinien des Kantons Bern, Modul 4: Sprache*. Justiz-, Gemeinde- und Kirchendirektion und Staatskanzlei des Kantons Bern, Bern.
- P. Clark, P. Harrison, T. Jenkins, J. Thompson, and R.H. Wojcik. 2005. Acquiring and using world knowledge using a restricted subset of English. In *FLAIRS 2005*, pages 506–511.
- B. Cohen. 2001. On the generic use of indefinite singulars. *Journal of Semantics*, 18(3):183–209.
- N.E. Fuchs, K. Kaljurand, and T. Kuhn. 2008. Attempto Controlled English for knowledge representation. In C. Baroglio, P.A. Bonatti, J. Maluszynski, M. Marchiori, A. Polleres, and S. Schaffert, editors, *Reasoning Web: 4th International Summer School 2008*, pages 104–124, Berlin. Springer.
- L. T. F. Gamut. 1991. *Logic, Language, and Meaning*, vol-ume 1. The University of Chicago Press, Chicago.
- A. Gardner. 1987. *An Artificial Intelligence Approach to Legal Reasoning*. MIT Press, Cambridge, MA.
- S. Hoefler and A. Bünzli. 2010. Controlled Legal German 1.0: Einführung und Spezifikation. Technical report, De-partment of Informatics, University of Zurich, Zurich.
- M. Hoey and C. Walter. 1988. Natural language interfaces. In C. Walter, editor, *Computer Power and Legal Lan-guage*, pages 135–142, New York. Quorum.
- B. Macias and S.G. Pulman. 1995. A method for control-ling the production of specifications in natural language. *The Computer Journal*, 38(4):310–318.
- L. T. McCarty. 2007. Deep semantic interpretation of legal texts. In *Proceedings of the 11th International Confer-ence on Artificial Intelligence and Law*, pages 217–224, New York. ACM Press.
- D. Mollá. 2001. Ontologically promiscuous flat logical forms for NLP. In *Proceedings of the 4th International Workshop on Computational Semantics (IWCS-4)*, pages 249–265, Tilburg.
- S. Neumann. 2009. Improving the comprehensibility of German court decisions. In G. Grewendorf and M. Rathert, editors, *Formal Linguistics and Law*. Moun-ton de Gruyter, Berlin.
- M. Nussbaumer. 2005. Zwischen Rechtsgrundsätzen und Formularsammlung: Gesetze brauchen (gute) Vagheit zum Atmen. In V. K. Bhatia, J. Engberg, M. Gotti, and D. Helier, editors, *Vagueness in Normative Texts*, pages 49–71. Peter Lang, Bern.
- G. J. Pace and M. Rosner. 2009. A controlled language for the specification of contracts. In N.E. Fuchs, editor, *Pre-proceedings of the Workshop on Controlled Natural Language (CNL 2009)*. CEUR-WS.
- J. Pool. 2006. Can controlled languages scale to the web? In *5th International Workshop on Controlled Language Applications*.
- E. L. Rissland, K. D. Ashley, and R. P. Loui. 2003. AI and law: A fruitful synergy. *Artificial Intelligence*, 150(1–2):1–15.
- U. Schwertel. 2000. Controlling plural ambiguities in At-tempto Controlled English. In *Proceedings of the 3rd In-ternational Workshop on Controlled Language Applica-tions*, Seattle, Washington.
- R. Schwitter and M. Tilbrook. 2006. Let’s talk in descrip-tion logic via controlled natural language. In *Proceed-ings of the 3rd International Workshop on Logic and En-gineering of Natural Language Semantics*, pages 193–207, Tokyo.
- R. Schwitter, A. Ljungberg, and D. Hood. 2003. ECOLE: A look-ahead editor for a controlled language. In *Controlled Translation, Proceedings of EAMT-CLAW03*, pages 141–150, Dublin.
- S. Spreeuwenberg and K. Anderson Healy. 2009. SBVR’s approach to controlled natural language. In N.E. Fuchs, editor, *Pre-Proceedings of the Workshop on Controlled Natural Language (CNL 2009)*. CEUR-WS.
- R. C. Wydick. 2005. *Plain English for Lawyers*. Carolina Academic Press, 5. edition.
- ZH (Regierungsrat des Kantons Zürich), Zürich, 2005. *Richtlinien der Rechtsetzung*.