

Can Controlled Languages Scale to the Web?

Jonathan Pool

Turing Center
University of Washington
Box 352350
Seattle, Washington 98195-2350, U.S.A.
<http://www.turing.washington.edu>
pool@cs.washington.edu
(206) 221-7701

Utilika Foundation
WRF Venture Center
2815 Eastlake Avenue East, Suite 300
Seattle, Washington 98102-3086, U.S.A.
<http://utilika.org>
pool@utilika.org
(206) 336-5566

Abstract

In a multilingual Semantic Web, authors might write in precise, expressive varieties of diverse languages. Do such controlled languages exist? Of 41 candidates, just 4 were (1) designed for multiple domains and genres and (2) documented enough for evaluation. A sample of Web statements on health and human rights revealed limited expressivity or precision in each language. The most expressive one avoided structural ambiguity but allowed semantic ambiguity that could frustrate human and machine comprehension. The possibility of a practical Web-scale controlled language remains undemonstrated but unrefuted.

Introduction

Of various Semantic Web visions (Marshall 2003), the most prominent (Berners-Lee 2001) imagines authors using "off-the-shelf software for writing Semantic Web pages" that machines can reason with. So that authors need not be knowledge engineers (Marshall 2003, Shirky 2003), formal but "seemingly informal" controlled natural languages might make semantic precision practical for them (Bernth 1998a; Clark 2005; Schwitter 2005). If so, equivalent controlled varieties of all languages could make the Semantic Web panlingual.

To evaluate the costs, benefits, and feasibility of a controlled-language Semantic Web, including the usability of controlled languages for humans and their tractability for machines, we need appropriate languages. Their documentation must show authors how to represent diverse illocutionary forces, evidentialities, probabilities, times, aspects, moods, numbers, persons, discourse references, entities, and relations. Unlike natural languages, however, such controlled languages must not merely permit but require authors to avoid structural and semantic ambiguities that frustrate automated natural-language processing.

Is any controlled natural language Web-ready? I considered projects in the last 25 years, whether their languages were aimed at the Web, machine reasoning, machine translation, or human intelligibility. I found 41 attempts to define (written) controlled varieties of English, Esperanto, French, German, Greek, Japanese, Mandarin, Spanish, or Swedish. A "controlled variety of X" licenses some but not all sentences of X, may require annotations, and may license sentences only resembling X. It may be *formalistic* (a language-like formal notation) or *naturalistic* (a language with restrictions), roughly equivalent to the "machine-oriented"/"human-oriented" distinction (O'Brien 2003, p. 1; Reuther 2003; Schwitter 2006, p. 2).

I did not consider controlled editing systems (e.g., Power 2004), natural-language-like programming languages (e.g., Apple 1999), and (3) natural-language-based designs for universal, philosophical, and exploratory languages (e.g., Harrison 2002, Langmaker 2006).

Screening

In Exhibit 1, I describe each project and classify it as "restrictive" or "general". Restrictive projects (22) overtly or apparently aim for expressivity in a domain (e.g., truck repair) and/or genre (e.g., instructions) and do not specify how to extend this expressivity. General projects (19) aim at languages for multiple domains and genres. I found 4 general, multidomain-multigenre languages sufficiently documented for evaluation.

Evaluation Procedure

My evaluation was exploratory. I selected two test domains: health information and international human rights. From Web sites (in English) in these domains, I chose ambiguous sentences, aiming to discover limits in precision with a small sample. I then attempted to translate each sentence into each controlled language, following the examples and instructions in its documentation, as any author might do. The sample sentences are:

1. Avoid prolonged exposure to excessive heat and humidity. (NLM 2005, art. 3217)
2. Mosquitoes have become resistant to the pyrethroid insecticide used to treat mosquito netting. (NIAID 2002, p. 12)
3. Scientists do not think this is a serious limitation yet. (NIAID 2002, p. 12)
4. The investigators found that the incidence of cancers of the nervous system and the blood was roughly 2.5 times higher in children whose mothers received pre-1963 vaccine than in children whose mothers did not. (NCI 2005)
5. Unless specific measures are taken to extend coverage and promote uptake in all population groups simultaneously, improvement of aggregate population coverage will go through a phase of increasing inequality. (WHO 2005, ch. 2, p. 30)
6. What type of illness do you suffer from most? (ERP 2005, q. 8)
7. Recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world. (UDHR 1948, Preamble)
8. Men and women of full age, without any limitation due to race, nationality or religion, have the right to marry and to found a family. (UDHR 1948, Art. 16)
9. No State Party shall expel, return ("refouler") or extradite a person to another State where there are substantial grounds for believing that he would be in danger of being subjected to torture. (CAT 1984, Art. 3)
10. The members of the Committee shall be elected by secret ballot from a list of persons nominated by States Parties. (CAT 1984, Art. 17)
11. Each State Party may nominate one person from among its own nationals. (CAT 1984, Art. 17)
12. An employer is required to take reasonable steps to accommodate your disability unless it would cause the employer undue hardship. (OCR n.d.)

The sentences are expressively diverse. They describe, identify, forecast, prescribe, recommend, declare, and ask. They make first-order (P), second-order (X believes P; X asserts P), and conditional (P if Q) assertions and absolute (do X) and conditional (do X if P) prescriptions. They deal with persons, animals, microorganisms, organizations, physical objects, substances, attributes, actions, states, events, concepts, and classes. They contain second- and third-person pronominal references. The entities in them are bare and quantified, definite and

indefinite, singular and plural, simple and modified. They describe various (simple, recent, continuing, and pre-past) pasts, various (recent, eternal, and temporary) presents, and the future. Acts include those with and without patient (object) arguments. Agents are specified and unspecified. There is conjunctive ("and") and disjunctive ("or") coordination.

The ambiguities are many (Pool 2006). Some are structural: coordination (1, 7), adverb-attachment (3, 5), prepositional-phrase-attachment (5, 7, 8), participle-attachment (10), clause-attachment (9). Others are semantic-pragmatic: joint/several (1, 4, 8), class/instance (2), restrictive/unrestrictive (2), perfective/imperfective (2), compound-relation (2), bare-plural-quantification (3), negation/inversion (3), negated-element (4), copular (7), word-sense (4, 5, 6, 7, 9, 12), presupposition/prescription (10), permission/prohibition (11), pronominal-reference (12), argument underspecification (1, 5), command/advice (1), assertive/effective/verdictive (8).

Consider sentence 1's ambiguities. Structural: The coordinand paired with "humidity" is any of these bracketed phrases: [prolonged [exposure to [excessive [heat]]]]. Pragmatic: The verb is imperative, but the illocutionary force may be a command or advice. Semantic: (1) Is one to avoid the two conditions severally or jointly? (2) The addressee (implicit subject of "avoid") may be the experiencer of "exposure" (as is plausible in health-care recommendations) or its agent (as in painting instructions).

Evaluations

I attempted to represent the test sentences in the 4 general, evaluatably documented languages and to determine whether the languages' rules require authors to prevent the ambiguities discovered in the sentences.

Formalized English (FE) (Martin 2002, Martin 2006) is being "designed to be as intuitive as possible" for English-like but deep-logic-based representations of "natural language sentences and knowledge ... in general". One meaning of sentence 1 in FE (version 2) might be "Any exposure with duration an important time and with object some humidity and with object some heat that is source of some pain should be object of an avoidance." One meaning of sentence 3 might be "No scientist is agent of a believing with object `*this is agent of a limitation with time *now'." And one of sentence 6's meanings might be "What is the illness that is the_most_frequent of the set of [illness that is supertype of an illness that has for object *you]?" (Words with "*" are not primitives and must be user-defined). As in these encodings, FE originally relied on the nominalization or nominal treatment of most verbs and adjectives, but phrases like "is agent of an extradition with object a person" now have alternates like "is extraditing a person". In lieu of tenses or utterance-relative time expressions, there is absolute time modification, such as "at time 2008". Conditionality can be represented with "if ... then", and possibility with "can", as in "Any state can be agent of a nomination" (from sentence 11). Various ambiguities are avoided in FE translation: "Fingers have joints" becomes "Any finger has for part at least 1 joint"; "women have rights" might become "any woman has for entitlement at least 1 right"; "Alex has malaria" might become "Alex has for infection malaria"; and "children have fathers" might become "any child has for father 1 person".

It is not clear, however, how to encode some sentences. The attempt in sentence 1 to give advice may instead have produced a judgment. Issuing commands (9, 10), granting permission (11), making second-person references (6, 12), expressing multiply-embedded comparisons (5), and stating presuppositions as in "not yet" (3) and "recognition" (7) are other potential expressive features of FE, but author guidance for them has not been published.

E2V (Pratt-Hartmann 2003) aims to be general-purpose and efficiently processable. Unlike Formalized English, E2V permits an unlimited set of verbs. "Every official who murders a

citizen deserves an imprisonment" and "Some patient does not protect herself" are E2V sentences. Pronouns and quantifiers can be used without ambiguity. For example, "Some dog which infects a monkey kills every cat which bites it" is ambiguous in English, but not in E2V, where "it" stands for "that dog" because the phrase-structurally nearest noun phrase that can be coreferential with nonreflexive "it" is the one headed by "dog", and the scopes of quantifiers "some", "a", and "every" are ordered by subject > object and NP > relative clause rules.

E2V's more naturalistic syntax may enhance its usability, but completing the evaluation would have required more documentation. It is not clear how to use E2V to give advice, issue commands, ask questions, specify times, state conditions, or express capability. Its designer acknowledges that it is not yet "a practically useful controlled language" (Pratt-Hartmann 2003, p. 14). Such broad applicability will depend on future extensions.

Attempto Controlled English (ACE) (Fuchs 2006) has been under development since 1995 as an intuitive but unambiguous fragment of English suitable for knowledge representation in the Semantic Web. As of version 4, its grammar licenses "countable and mass nouns, collective and distributive plurals, generalised quantifiers, indefinite pronouns, phrasal and prepositional verbs, phrasal and sentential negation, and anaphoric references to noun phrases through proper names, definite noun phrases, pronouns, and variables" (Fuchs 2006, p. 1). The associated parser can translate any valid ACE sentence into a first-order representation. Its original purpose of representing project specifications has been extended to include, for example, database integrity constraints, business rules, and protein interactions. Of the evaluated languages, ACE is the one under most active development.

Sentence 7 could be an ACE sentence if amended to "*The* recognition of the inherent dignity and *of* the equal and inalienable rights of *all every* members of the human family is the foundation of *the* freedom *and the* justice and *the* peace in the world." The coordination ambiguities would be resolved as apparently intended (but see Pool 2006), though other sentences, such as "The woman is the owner of a dog and the mother of two boys" (where the woman would be the owner of the mother), exhibit what can be called deceptive precision.

Nonetheless, the evaluation calls for some features not in ACE (Fuchs 2005), including intentionality, modality, and clausal complementation (all planned for version 5). This conflicts with sentences 3 ("think [that]"), 4 ("found that"), 5 ("taken [in order] to"), 8 ("right to"), 9 ("believing that"), 10 ("shall be"), 11 ("may"), and 12 ("is required to", "steps to"). "Verbs are restricted to simple present tense, third person singular and plural, active voice, and indicative mood" (Fuchs 2006, p. 2), ruling out sentences 1, 2, 4, 5, 6, and 9.

The *DLT Intermediate Language (DLTIL)* (Witkam 1983, Schubert 1986, Schubert 2004) has been unused since 1986, but until then was being developed as a machine-translation interlingua for multidomain "informative" texts. It contains four open word classes (verb, noun, adjective, adverb) and about 300 function words partitioned into two open (adjectives, adverbs) and six closed classes (pronouns, articles, numerals, prepositions, conjunctions, and interjections). Its grammar and lexicon are based on Esperanto. In practice, to compose a sentence in DLTIL, one can compose it in Esperanto and modify it as required. With Esperanto ranking about 24 among languages on the Web (Kilgarriff 2003, p. 7), DLTIL can be expected to have wide expressive coverage, but questionable precision. DLTIL's design as a translation interlingua, too, makes one expect limited precision: One design principle was to make DLTIL about as ambiguous lexically as typical natural languages, so as not to force unnecessary disambiguation effort in source-to-interlingua translation (Witkam 1983, pp. III.15-III.45).

As expected, the sentences are all translatable into DLTIL, but the language does not avoid all the significant ambiguities (Pool 2006). DLTIL typically avoids the original structural ambiguities and some, but not all, of the semantic ambiguities. It generally avoids ambiguities

related to the command/advice distinction, verb-negation semantics, the active/passive semantics of nominalized verbs, closed-class word senses, negation scope, coordination syntax, and the attachments of adjuncts, adverbs, prepositions, clauses, and participles. Ambiguities that tend to survive include those relating to the description/declaration distinction, individual versus aggregate change, the prohibitions implied by permissions, the implied scope of commands, the implicit controlled argument of nominalized verbs, the aspectual interpretation of the recent past, quantitative comparison, implied thematic roles, existential implications, descriptive versus restrictive modification, implicit quantification, coordination semantics, open-class word senses, long-distance dependencies, pronominal reference, and compound modification sense.

For example, DLTIL can represent sentence 1, and in doing so it prevents two of the four ambiguities described above. One representation (with the orthography, morphemes, morpheme gloss, and English translation) is:

Evit`u la daŭr`a`n en`ad`o`n en tro`a`j varm`o kaj humid`o.

Evit-u la daŭr-a-0-n en-ad-o-0-n en tro-a-j-0 varm-o-0-0 kaj humid-o-0-0.

Avoid-IMP the last-ADJ-SG-ACC in-ing-N-SG-ACC in too-ADJ-PL-NOM hot-N-SG-NOM and humid-N-SG-NOM

Avoid prolonged exposure to excessive (1) heat and (2) humidity.

DLTIL inflects adjectives for number, and this partly prevents the coordinand ambiguity: The adjectives are (singular, plural) when the left coordinand is "heat", and (plural, singular) when it is "exposure to excessive heat". But they are (singular, singular) in the other two cases. DLTIL further requires that a conjunction be prefixed with a mark for each leftward enlargement of the minimal left coordinand. Thus, for the three possible coordinands larger than "heat", the conjunction "kaj" ('and') becomes ".kaj", "·kaj", and "··kaj". This (semiredundantly) prevents the coordinand ambiguity. DLTIL also requires that the imperative mood carry command force (Witkam 1983, p. IV.44) and uses another construction for advice. But the documentation does not specify a rule interpreting a coordination as joint or several or a rule identifying an ambiguous implicit argument of a nominalized verb like "exposure".

The ambiguities permitted by DLTIL may promote intelligible translation, but seem likely to hinder Semantic Web functionalities like retrieval, question answering, and summarization.

Conclusion

Controlled natural languages that have been reported as successes have been mainly restrictive: designed for limited, intra-organization or intra-industry purposes. That they cover single domains and genres, with repetitive and trainable authors, facilitates their efficacy.

General controlled natural languages--designed for multidomain, multigenre meaning expression as in the Semantic Web--will be deployed under less auspicious conditions. My evaluation suggests that for such uses formalistic languages (like FE, E2V, and ACE) will exhibit high precision but limited expressivity, while naturalistic ones (like DLTIL) will be highly expressive but semi-precise. The two strategies might converge, but no project has bridged the gap yet, and it remains unknown how a controlled natural language can achieve precise, yet broadly expressive, meaning representation.

Until we have Web-scale controlled natural languages, we cannot (1) evaluate their human usability and machine tractability or (2) develop and evaluate support systems for their use in Semantic Web authorship. Such systems could (1) customize grammars and lexicons, (2) use statistical methods to rank analyses, (3) learn each author's usual word senses, (4) check

validity, (5) construct valid expressions by interviewing authors, (6) help authors estimate the return on their controlled-encoding investment, and (7) help native speakers create standard-compliant controlled languages.

The two strategies of controlled-language design might enliven the quest for a Semantic Web, whose mainstream planning is based on the formalistic RDF semantic model and notation (W3C 2004, Shadbolt 2006). A controlled-language approach would imply competing scenarios. In a formalistic scenario, the Semantic Web begins with precise but expressively confined annotations to existing content. In a naturalistic scenario, the Semantic Web begins with incremental disambiguating modifications to existing content. Controlled languages could thus help supplement the discussion of whether the Semantic Web vision is right or wrong with a discussion of which Semantic Web strategy is superior.

Exhibit 1: Controlled Natural Languages

Legend: "F/" = formalistic; "N/" = naturalistic; "/R" = restricted; "/G" = general; "Exp:" = Expressiveness restrictions; "Mod:" = notable modifications to base language's grammar; "Lex" = closed lexicon or limitations on lexeme senses.

Airbus Warning Language (Spaggiari 2003): N/R, English. Exp: Short industrial warnings. Mod: No "un-" with "not" sense; word-order restrictions. Lex.

ALCOGRAM (Adriaens 1992): N/R, English. Exp: Telecommunication technology. Mod: No determiner omission or introductory participial-adverbial clauses. Lex.

ASD Simplified Technical English (ASD 2005): N/R, English. Exp: Prescriptions and descriptions. Mod: No gerunds, present participles, complex tenses, conditional or subjunctive mood, passive voice, or compounds with 4+ elements; no omission of articles except on noninitial coordinands. Lex.

Atempto Controlled English (Fuchs 2006): F/G, English. Exp: Specifications. Mod: Dynamic names; no 1st- and 2nd-person pronouns, tenses, continuous aspect, passives, subjunctives, imperatives, modals, intensional verbs, copular "become", definite copula complements (but contrary example on p. 61), ditransitive verbs except with NP and "to"-PP as complements, non-nominal comparisons (e.g., "he speaks faster than he writes"), superlative predication, "by", "under", "over", "behind", "in front of", "before", or "after"; no adverbs modifying sentences, adjectives, or adverbs; no prepositional phrases modifying sentences, adjectives, or (except "of") nouns; no coordination of subjects, prepositional objects, or "of" phrases. Lex.

Avaya Controlled English (O'Brien 2003): N/R, English.

CELT (Pease 2003): F/R, English. Lex: English WordNet mapped to SUMO; polysemy permitted.

ClearTalk (Skuce 2003): F/R, English. Exp: Prescriptions, descriptions. Mod: Parentheses must resolve ambiguous attachments and coordinations; verb conjunctions analyzed as having no sequential denotation.

CLIP (Sukkarieh 2003a, Sukkarieh 2003b): F/R, English. Exp: McLogic-equivalent sentences. Mod: Quantifier scopes must decrease in left-to-right surface order; no coordination with collective (vs. distributive) sense.

Common Logic Controlled English (Sowa 2004): F/R, English. Exp: First-order-logic-equivalent sentences. Mod: Nouns must be singular; verbs must be present-tense; noncompositional nominals must be declared as units; no pronouns; restrictions on prepositional and quantifier attachment; annotation requirements.

Controlled Automotive Service Language (Hebling 2002, pp. 107-112): N/R, English. Exp: Automotive maintenance. Mod: No personal pronouns; 62 restrictive rules. Lex.

Controlled Chinese (Zhang 1998): N/G, Mandarin. Lex: No polysemy.

Controlled English (Océ) (Cucchiaroni 2002, p. 2; Cremers 2003, pp. 38-51): N/R, English. Exp: Prescriptions and descriptions. Mod: as with ASD Simplified Technical English. Lex.

Controlled Modern Greek (Vassiliou 2003): N/G, Greek. Mod: Some constructions prohibited. Lex.

CPL (Clark 2005, Clark 2006): F/G, English. Exp: Knowledge-Machine-equivalent sentences, including nonpolar questions. Mod: No pronouns; structural restrictions.

DLT Intermediate Language (Witkam 1983; Schubert 1986; Schubert 2004): N/G, Esperanto. Exp: Nonmetaphorical statements. Mod: Intraword morpheme boundaries, prepositional-phrase attachment distances, and coordination ellipsis must be marked. Lex.

E2V (Pratt-Hartmann 2003): F/G, English. Exp: Two-variable first-order-logic-equivalent sentences. Mod: No long-distance pronominal reference; universal, subject, and main-clause quantifiers must have wide scope; scope must obey subject quantifier > clausal negation > quantifier within negated clause.

EasyEnglish (Betts 2003): N/G, English. Mod: No pronouns with ambiguous antecedents; no ambiguous genitive case. Lex.

EasyEnglishAnalyzer (Bernth 1997; Bernth 1998a; Bernth 1998b; Hebling 2002, pp. 93-106): N/G, English. Mod: Post-complement participial phrases deprecated; no shared constituent coordination; noun-participle modifiers must be hyphenated; embedded coordinands must be bracketed; at most 1 embedded coordination must be nonbinary; no preclausal modifiers of non-subject arguments; no pronouns with ambiguous antecedents; no postmodified double passive participles; 40 restrictive or deprecative rules.

Ericsson English (Adriaens 1992): N/R, English. Mod: No present participles. Lex: Closed lexicon.

FAA Air Traffic Control Phraseology (FAA 2006, Jones 2002): N/R, English. Exp: Air traffic control.

First Order English (Pulman 2002): F/G, English. Exp: First-order-logic-equivalent sentences.

Formalized English (Martin 2002, Martin 2006): F/G, English. Exp: Conceptual-Graph-Interchange-Format- and thus first-order-logic-equivalent sentences. Mod: No imperative mood or first- or second-person pronouns.

Français rationalisé (Barthe 1999): N/R, French. Exp: Prescriptions and descriptions. Mod: Noun compounds with omitted prepositions deprecated; no past infinitives (e.g., "avoir aidé"), prepositional phrases of time ("lors de"), subjunctive verbs, or unregistered reflexive verbs; no future tense with present or imperative sense; 50 restrictive rules. Lex: Lexical restrictions include exclusion of some verb frames (e.g., "empêcher ... de [VINFIN]").

interNOSTRUM Controlled Spanish (Canals-Marote 2001, p. 3): N/G, Spanish. Not implemented.

KANT Controlled English (Mitamura 1995; Mitamura 1999; Nyberg 1996): N/G, English. Exp: Concise technical prescriptions, descriptions, and questions. Mod: No implicit heads (Mitamura 1995, pp. 7-8), ellipsis (Mitamura 1995, p. 8), or object or complex relative clauses; prepositional-phrase attachment must comply with syntactic-semantic mapping rules and domain-model frames; SGML attachment tags permitted for attachment disambiguation (Mitamura 1995, p. 9; Nyberg 1996, p. 7); coordination must comply with unique nondistributed interpretation rule (Nyberg 1996, pp. 7-8); pronoun antecedents must comply with discourse reference resolution rules (Mitamura 2002); some compound-noun modification senses

prohibited (Mitamura 1995, p. 7). Lex.

MenuChoice (Vertan 2003): F/R, English. Exp: Tourist health conversations.

MULTILINT (Reuther 1998): N/G, German. Mod: Genitive arguments of verb nominalizations with ambiguous thematic roles and some categorial and other structural ambiguities deprecated.

Multinational Customized English (Adams 1999): N/G, English. Lex: No polysemy.

PENG (Schwitter 2003, Schwitter 2006): F/R, English. Exp: Specifications and use cases equivalent to first-order-predicate-logic sentences and questions answered by them.

Perkins Approved Clear English (Douglas 1996): N/G, English. Mod: 10 restrictive rules. Lex.

Plain Japanese (Sato 2004): N/G, Japanese. Mod: No topic-marking or idiomatic use of "nara". Lex: closed lexicon, spelling restrictions.

PoliceSpeak (Johnson 2002): N/R, English. Exp: Police radio communication. Lex.

ScaniaSwedish (Almqvist 1996, Uppsala 2001): N/R, Swedish. Exp: Automotive technology. Lex.

SeaSpeak (Kimbrough 2004): N/R, English. Exp: Maritime communication. Mod: Illocutionary force must be marked. Lex.

Siemens-Dokumentationsdeutsch (Lehrndorfer 1998): N/G, German.

Simplified Technical Spanish (Ruiz 2003): N/R, Spanish. Exp: Prescriptions and descriptions. Mod: approximately as with ASD Simplified Technical English.

Simplus (Lingua 2006): N/G, English. Mod: 60 restrictive rules, partly based on ASD Simplified Technical English. Lex: user-definable lexicons.

Sun Proof (Akis 2003, O'Brien 2003): N/R, English. Mod: No future tense, 3rd-person pronouns, reduced relative clauses, gerunds, present participles, or clusters of 4+ nouns.

TITUS (Lehtola 1999a): N/R, French. Exp: Textiles.

Universal Translation Language (Franco 2001): N/G, Esperanto. Mod: Relative and interrogative proforms and personal and impersonal senses of u-series correlatives lexically differentiated; nominal attachment of prepositional phrases must be marked with preposition preposition; pronominal antecedents must be marked if ambiguous; no idiomatic expressions. Lex: No polysemy.

Webtran (Lehtola 1999b): N/R, Swedish. Exp: Product catalogs. Mod: Restrictive rules user-specifiable. Lex: Lexicon user-specifiable.

References

- Adams 1999. Ann H. Adams, Gail W. Austin, and Melissa Taylor, "Developing a Resource for Multinational Writing at Xerox Corporation", *Technical Communication*, 46, 1999, 249-254. <http://www.ingentaconnect.com/content/stc/tc/1999/00000046/00000002/art00012>, <http://64.233.179.104/search?q=cache:mzQzuFE2kXEJ:www.plainlanguage.gov/hotstuff/xerox.htm>
- Adriaens 1992. Geert Adriaens and Dirk Schreurs, "From COGRAM to ALCOGRAM: Toward a Controlled English Grammar Checker", *COLING-92*, 1992. <http://acl.ldc.upenn.edu/C/C92/C92-2090.pdf>
- Akis 2003. Jennifer Wells Akis, Stephanie Brucker, Virginia Chapman, Layne Ethington, Bob Kuhns, and PJ Schemenaur, "Authoring Translation-Ready Documents: Is Software the Answer?", *SIGDOC '03*, 2003. <http://portal.acm.org/citation.cfm?id=944878>
- Almqvist 1996. Ingrid Almqvist and Anna Sagvall Hein, "Defining ScaniaSwedish--A Controlled Language for Truck Maintenance", ms, 1996. <http://www.lingfil.uu.se/personal/anna/claw1.pdf>

- Apple 1999. Apple Computer, Inc., *AppleScript Language Guide*, 1999. <http://developer.apple.com/documentation/AppleScript/Conceptual/AppleScriptLangGuide/AppleScriptLanguageGuide.pdf>
- ASD 2005. AeroSpace and Defence Industries Association of Europe, *ASD Simplified Technical English: Specification ASD-STE100*, 2005. http://www.simplifiedenglish-aecma.org/Simplified_English.htm
- Barthe 1999. Kathy Barthe, Claire Juaneda, Dominique Leseigneur, Jean-Claude Loquet, Claude Morin, Jean Escande, and Annick Vayrette, "GIFAS Rationalized French: A Controlled Language for Aerospace Documentation in French", *Technical Communication*, 46, 1999, pp. 220-229. <http://www.ingentaconnect.com/content/stc/tc/1999/00000046/00000002/art00009>
- Berners-Lee 2001. Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web", *Scientific American*, 284(5), 2001, 34-43. <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>
- Bernth 1997. Arendse Bernth, "EasyEnglish: A Tool for Improving Document Quality", *ANLP 1997*, 1997. <http://acl.ldc.upenn.edu/A/A97/A97-1024.pdf>
- Bernth 1998a. Arendse Bernth, "EasyEnglish: Preprocessing for MT", *CLAW '98*, 1998.
- Bernth 1998b. Arendse Bernth, "EasyEnglish: Addressing Structural Ambiguity", *AMTA '98*, 1998. <http://www.springerlink.com/link.asp?id=8kpl65y9hl6rgvt2>
- Betts 2003. Robert G. Betts, "Wycliffe Associates EasyEnglish: Challenges in Cross-Cultural Communication", *EAMT-CLAW 2003*, 2003. <http://www.mt-archive.info/CLT-2003-Betts.pdf>
- Canals-Marote 2001. R. Canals-Marote, A. Esteve-Guillén, A. Garrido-Alenda, M. I. Guardiola-Savall, A. Iturraspe-Bellver, S. Montserrat-Buendia, S. Ortiz-Rojas, H. Pastor-Pina, P.M. Pérez-Antón, and M. L. Forcada, "The Spanish<->Catalan Machine Translation System interNOSTRUM", ms, 2001. <http://www.internostrum.com/docum/iN-MTS.pdf>
- CAT 1984. Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, 1984. <http://www.ohchr.org/english/law/cat.htm>
- Clark 2005. Peter Clark, Phil Harrison, Tom Jenkins, John Thompson, and Rick Wojcik, "Acquiring and Using World Knowledge Using a Restricted Subset of English", *FLAIRS 2005*, 2005. <http://www.cs.utexas.edu/users/pclark/papers/flairs.pdf>
- Clark 2006. Peter Clark, Phil Harrison, John Thompson, Rick Wojcik, Tom Jenkins, and David Israel, "Reading to Learn: Final Report", ms, 2006. <http://www.cs.utexas.edu/users/pclark/rtol/final-report.doc>
- Cremers 2003. Lou Cremers, "Controlled Language in an Automated Localisation Environment", *EAMT-CLAW 2003*, 2003. <http://www.ctts.dcu.ie/Cremers.PPT>
- Cucchiariini 2002. Catia Cucchiariini, "Euromap HLT Case Study: How HLT Applications Can Lead to Higher Quality Translations at Lower Costs: The Experience of Océ Technologies", ms, 2002. http://www.hltcentral.org/usr_docs/case_studies/euromap/NL_ocecasestudy_nl.pdf
- Douglas 1996. Shona Douglas and Matthew Hurst, "Controlled Language Support for Perkins Approved Clear English (PACE)", ms, 1996. <http://www.ltg.ed.ac.uk/papers/claw96.ps>
- ERP 2005. Exercise Research Project, "Exercise and Health Questionnaire", ms, 2005. <http://salmon.psy.plym.ac.uk/Yr2Skills/clara.htm>
- FAA 2006. United States Federal Aviation Administration, Order 7110.65R, "Air Traffic Control", 2006. <http://www.faa.gov/ATpubs/ATC/INDEX.HTM>
- Franco 2001. Marcos Franco Sabarís, Jose Luis Rojas Alonso, Carlos Dafonte, and Bernardino Arcay, "Multilingual Authoring through an Artificial Language", ms, 2001. <http://www.eamt.org/summitVIII/papers/franco.pdf>

- Fuchs 2005. Norbert E. Fuchs, "Attempto Controlled English", ms, 2005. <http://www.ifi.unizh.ch/attempto/talks/files/Talk.Stanford.05.pdf>
- Fuchs 2006. Norbert E. Fuchs, Kaarel Kaljurand, and Gerold Schneider, "Attempto Controlled English Meets the Challenges of Knowledge Representation, Reasoning, Interoperability and User Interfaces", *FLAIRS 2006*, 2006. <http://www.ifi.unizh.ch/attempto/publications/papers/FLAIRS0601FuchsN.pdf>
- Harrison 2002. Richard K. Harrison, "Bibliography of Planned Languages (Excluding Esperanto)", ms, 2002. <http://www.rickharrison.com/language/bibliography.html>
- Hebling 2002. Uta Hebling, *Controlled Language am Beispiel des Controlled English*, Trier, Germany: Wissenschaftlicher Verlag Trier, 2002. <http://www.lighthouse-unlimited.de/shop/ak42414e44203034.htm>
- Johnson 2002. Edward Johnson, "Talking across Frontiers: Building Communication between Emergency Services", *Regional & Federal Studies*, 12, 2002, 88-110. <http://www.prolingua.co.uk/talking.pdf>
- Jones 2002. Kent Jones, "Misfunctional FAA Phraseology", ms, 2002. http://www.esperanto-sat.info/imprimersans.php3?id_article=351
- Kilgarriff 2003. Adam Kilgarriff and Gregory Grefenstette, "Web as Corpus", *Computational Linguistics*, 29 (3), 2003, 333-347. <http://www.kilgarriff.co.uk/Publications/2003-KilgGrefenstette-WACIntro.pdf>
- Kimbrough 2004. Steven O. Kimbrough, Thomas Y. Lee, Balaji Padmanabhan, and Yinghui Yang, "On Original Generation of Structure in Legal Documents", ms, 2004. <http://grace.wharton.upenn.edu/~sok/sokpapers/2004/kimbrough/kimbrough.pdf>
- Langmaker 2006. *Langmaker*, ms. http://www.langmaker.com/db/Main_Page
- Lehrndorfer 1998. Anne Lehrndorfer and Stefanie Schachtl, "TR09: Controlled Siemens Documentary German and TopTrans", *TC Forum*, 1998:3. <http://www.tc-forum.org/topic/tr9contr.htm>
- Lehtola 1999a. Aarno Lehtola, Jarno Tenni, Catherine Bounsaythip, and Kristiina Jaaranen, "Controlled Languages as the Basis for Multilingual Catalogues on the WWW", in Jean-Yves Roger, Brian Stanford-Smith, and Paul T. Kidd (eds.), *Business and Work in the Information Society: New Technologies and Applications*, Amsterdam: IOS-Press, 1999, pp. 207-213. <http://www.kolumbus.fi/lehtola.net/cache/emmsec99.pdf>
- Lehtola 1999b. Aarno Lehtola, Jarno Tenni, Catherine Bounsaythip, and Kristiina Jaaranen, "WEBTRAN: A Controlled Language Machine Translation System for Building Multilingual Services on Internet", ms, 1999. <http://www.kolumbus.fi/lehtola.net/cache/mtsummit1999.pdf>
- Lingua 2006. Lingua Technologies, Inc., "Simplus Data Sheet", ms, 2006. <http://www.linguatechnologies.com/simplus/english/SIMPLUSDataSheet.pdf>
- Marshall 2003. Catherine C. Marshall and Frank M. Shipman, "Which Semantic Web?", *Hypertext '03 Proceedings*, 2003. <http://www.csdl.tamu.edu/~marshall/ht03-sw-4.pdf>
- Martin 2002. Philippe Martin, "Knowledge representation in CGLF, CGIF, KIF, Frame-CG and Formalized-English", ms, 2002. <http://www.webkb.org/doc/papers/iccs02/iccs02.pdf>
- Martin 2006. Philippe Martin, "Formalised English", ms, 2006. <http://www.webkb.org/doc/languages/FE.html>
- Mitamura 1995. Teruko Mitamura and Eric H. Nyberg, 3rd, "Controlled English for Knowledge-Based MT: Experience with the KANT System", *TMI 95*, 1995. <http://www.lti.cs.cmu.edu/Research/Kant/PDF/tmi-95-final.pdf>
- Mitamura 1999. Teruko Mitamura, "Controlled Language for Multilingual Machine Translation", *MT Summit '99*, 1999. <http://www.lti.cs.cmu.edu/Research/Kant/PDF/MTSummit99.pdf>

- Mitamura 2002. Teruko Mitamura, Eric Nyberg, Enrique Torrejon, Dave Svoboda, Annelen Brunner, and Kathryn Baker, "Pronominal Anaphora Resolution in the KANTOO Multilingual Machine Translation System", *TMI 02*, 2002. <http://www.lti.cs.cmu.edu/Research/Kant/PDF/tmi02-anaphora-camera.pdf>
- NCI 2005. National Cancer Institute, "Studies Find No Evidence That SV40 is Related to Human Cancer", ms, 2005. <http://www.cancer.gov/newscenter/pressreleases/SV40>
- NIAID 2002. National Institute of Allergy and Infectious Diseases, *Malaria*, 2002. <http://www.niaid.nih.gov/publications/malaria/pdf/malaria.pdf>
- NLM 2005. National Library of Medicine, *Medical Encyclopedia*, 2005. <http://www.nlm.nih.gov/medlineplus/encyclopedia.html>
- Nyberg 1996. Eric H. Nyberg, 3rd, and Teruko Mitamura, "Controlled Language and Knowledge-Based Machine Translation: Principles and Practice", *CLAW '96*, 1996. <http://www.lti.cs.cmu.edu/Research/Kant/PDF/claw.pdf>
- O'Brien 2003. Sharon O'Brien, "Controlling Controlled English: An Analysis of Several Controlled Language Rule Sets", *EAMT-CLAW 2003*, 2003. <http://www.mt-archive.info/CLT-2003-Obrien.pdf>
- OCR n.d. United States Department of Health and Human Services, Office for Civil Rights, "Your Rights under Section 504 of the Rehabilitation Act", n.d.. <http://www.hhs.gov/ocr/504.html>
- Pease 2003. Adam Pease and William Murray, "An English to Logic Translator for Ontology-Based Knowledge Representation Languages", *NLPKE 2003*, 2003. <http://home.earthlink.net/~adampease/professional/Pease-NLPKE.pdf>
- Pool 2006. Jonathan Pool, "Can Controlled Languages Scale to the Web?: Evaluation of the DLT Intermediate Language", ms, 2006. <http://utilika.org/pubs/etc/ambigcl/evdlt.html>
- Power 2004. Richard Power and Roger Evans, "WYSIWYM with Wider Coverage", *ACL 2004*, 2004. <http://www.itri.brighton.ac.uk/~Richard.Power/acl04.pdf>
- Pratt-Hartmann 2003. Ian Pratt-Hartmann, "A Two-Variable Fragment of English", *Journal of Logic, Language and Information*, 12, 2003, 13-45. http://www.cs.man.ac.uk/~ipratt/papers/nat_lang/e2v.ps
- Pulman 2002. Stephen G. Pulman, "Current Projects in Computational Linguistics", ms, 2002. http://www.clp.ox.ac.uk/people/staff/pulman/current_projects.html
- Reuther 1998. Ursula Reuther, "Controlling Language in an Industrial Application", ms, 1998. <http://www.iai.uni-sb.de/docs/clrev.pdf>
- Reuther 2003. Ursula Reuther, "Two in One--Can it work? Readability and Translatability by means of Controlled Language", *EAMT-CLAW 2003*, 2003. <http://www.mt-archive.info/CLT-2003-Reuther.pdf>
- Ruiz 2003. Remedios Ruiz Cascales and Richard F.E. Sutcliffe, "A Specification and Validating Parser for Simplified Technical Spanish", *EAMT-CLAW 2003*. <http://www.mt-archive.info/CLT-2003-Ruiz.pdf>
- Sale 1993. Sale v. Haitian Ctrs. Council, 113 S. Ct. 2549, 125 L. (92-344), 509 U.S. 155 (1993). <http://straylight.law.cornell.edu/supct/html/92-344.ZS.html>
- Sato 2004. Satoshi Sato, Takehito Utsuro, Masatoshi Tsuchiya, Masahiro Asaoka, and Suguru Matsuyoshi, "Natural Language Processing Technologies to Enhance Readability", *ICKS 2004*, 2004. http://ieeexplore.ieee.org/xpl/abs_free.jsp?arNumber=1313408
- Schubert 1986. Klaus Schubert, *Syntactic Tree Structures in DLT*, Utrecht, Netherlands: BSO/Research, 1986.
- Schubert 2004. Klaus Schubert, "Projekt Distributed Language Translation", ms, 2004. <http://www.wi.fh-flensburg.de/ifk/schubert/forschung/FuEDLTInh.htm>

- Schwitter 2003. Rolf Schwitter, "PENG in a Nutshell", ms, 2003. <http://www.ics.mq.edu.au/~rolfs/peng/nutshell.html>
- Schwitter 2005. Rolf Schwitter, "Controlled Natural Language as Interface Language to the Semantic Web", *IICAI-05*, pp. 1699-1718, 2005. <http://www.ics.mq.edu.au/~rolfs/papers/IICAI-schwitter-2005.pdf>
- Schwitter 2006. Rolf Schwitter and Marc Tilbrook, "Writing RSS Feeds in a Machine-Processable Controlled Natural Language", *CLAW 2006*, 2006. <http://www.ics.mq.edu.au/~rolfs/>
- Shadbolt 2006. Nigel Shadbolt, Wendy Hall, and Tim Berners-Lee, "The Semantic Web Revisited", *IEEE Intelligent Systems*, 21, 2006, 96-101. http://eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisited.pdf
- Shirky 2003. Clay Shirky, "The Semantic Web, Syllogism, and Worldview", ms, 2003. http://www.shirky.com/writings/semantic_syllogism.html
- Skuce 2003. Doug Skuce, "A Controlled Language for Knowledge Formulation on the Semantic Web", ms, 2003. <http://www.site.uottawa.ca:4321/factguru2.pdf>
- Sowa 2004. John F. Sowa, "Common Logic Controlled English", ms, 2004. <http://www.jfsowa.com/clce/specs.htm>
- Spaggiari 2003. Laurent Spaggiari, Florence Beaujard, and Emmanuelle Cannesson, "A Controlled Language at Airbus", *EAMT-CLAW 2003*. <http://www.mt-archive.info/CLT-2003-Spaggiari.pdf>
- Sukkarieh 2003a. Jana Z. Sukkarieh, "An Expressive Efficient Representation: Bridging a Gap between Natural Language Processing (NLP) and Knowledge Representation (KR)", *KES 2003*, 2003. http://www.ling-phil.ox.ac.uk/people/staff/jana/Bridging_gap03.pdf
- Sukkarieh 2003b. Jana Z. Sukkarieh, "Mind your Language! Controlled Language for Inference Purposes", *EAMT-CLAW 2003*, 2003. <http://www.mt-archive.info/CLT-2003-Sukkarieh.pdf>
- UDHR 1948. Universal Declaration of Human Rights, 1948. <http://www.unhchr.ch/udhr/navigate/alpha.htm>
- Uppsala 2001. Uppsala Universitet, "The Scania Project", ms, 2001. <http://stp.lingfil.uu.se/~corpora/scania/>
- Vassiliou 2003. Marina Vassiliou, Stella Markantonatou, Yanis Maistros, and Vangelis Karkaletsis, "Evaluating Specifications for Controlled Greek", *EAMT-CLAW 2003*, 2003. <http://www.mt-archive.info/CLT-2003-Vassiliou.pdf>
- Vertan 2003. Cristina Vertan and Walther v. Hahn, "Menu Choice Translation: A Flexible Menu-Based Controlled Natural Language System", *EAMT-CLAW 2003*, 2003. <http://www.mt-archive.info/CLT-2003-Vertan.pdf>
- W3C 2004. World Wide Web Consortium, *RDF Primer*, 2004. <http://www.w3.org/TR/rdf-primer/>
- WHO 2005. World Health Organization, *The World Health Report 2005: Make Every Mother and Child Count*, 2005. http://www.who.int/entity/whr/2005/whr2005_en.pdf
- Witkam 1983. A.P.M. Witkam, *Distributed Language Translation: Feasibility Study of a Multilingual Facility for Videotex Information Networks*, Utrecht, Netherlands: BSO, 1983.
- Zhang 1998. Zhang Wei, Zhou Xiling, and Yu Shiwen, "Construction of Controlled Chinese Lexicon", 1998. <http://scholar.google.com/scholar?q=cache:eFl91MyxwpQJ:www4.ncsu.edu/~wzhang/zwclaw98.pdf>

I acknowledge helpful extensive comments on prior drafts from Emily Bender, Philippe Martin, Norbert Fuchs, Susan Colowick, and anonymous reviewers.