

Controlled Language

Referat von Tobias Schweizer
Juni 2004

Inhalt

- Was ist eine Controlled Language (CL)?
- Mehrdeutigkeit natürlicher Sprachen
- Reduktion von Mehrdeutigkeit durch CL
- Einsatzmöglichkeiten von CL
- Simplified English (SE)
- Nutzen von CL

Definition

Eine CL ist die gezielte Einschränkung einer natürlichen Sprache wie beispielsweise Englisch. Dabei bestehen Einschränkungen in Bezug auf das Lexikon (Wortschatz), die Grammatik und den Stil.

Probleme mit natürlicher Sprache

- homonymy
- synonymy
- conversions
- inkonsistente Schreibweise von Worten

Lexikon einer CL

- Kleines, kontrolliertes Vokabular
- eine Bedeutung pro Wort (falls möglich)
- Aufteilung von Bedeutung mit Synonymen
- keine Homonyme
- keine Synonyme
- keine Konversionen
- konsistente Schreibweise von Worten

Lexikoneinträge

Worte sind entweder zulässig (approved) oder unzulässig (unapproved). Ist ein Wort nicht zulässig, so kann seine Bedeutung mit einer zulässigen Alternative ausgedrückt werden. Die Möglichkeiten des sprachlichen Ausdrucks können dadurch eingeschränkt werden.

Mehrdeutigkeit eines Worts

Das englische Wort *right* kann sowohl richtig als auch rechts bedeuten. Im Lexikon einer CL hat das Wort *right* jedoch nur eine Bedeutung, z.B. rechts. Die zweite Bedeutung muss durch ein Synonym ausgedrückt werden wie *correct*. Die Bedeutung des Wortes *right* ist somit für Mensch und Maschine eindeutig.

Grammatikalische Zweifelsfälle

Bezüge von Präpositionen können zweideutig sein. Für Menschen stellt das im Gegensatz zu Computern oft kein Problem dar.

Beispiel:

piece of glass and metal

1. Möglichkeit: *piece of [glass and metal]*
2. Möglichkeit: *[piece of glass] and [metal]*

Grammatik einer CL

- klare Bezüge von Präpositionen
- Teilsätze sollten vom gleichen Typ sein
- Relativsätze: Relativpronomen sind das Subjekt
- keine Satzellipsis
- beschränkte Satzlänge (z.B. max. 20 Worte)
- Passivform nicht verwenden

Ziel einer CL

Reduktion der Ambiguität (Mehrdeutigkeit) und Komplexität sowohl auf der lexikalischen als auch auf der grammatikalischen Ebene.

Einsatz von CL

Es wird zwischen human-oriented (z.B. technische Dokumentationen) und machine-oriented (MT) CL unterschieden. Texte werden für Menschen und Maschinen leichter verständlich gemacht. Regeln für machine-oriented CL müssen jedoch maschinell verarbeitbar sein, während Menschen auch vague formulierte Anweisungen verstehen.

Nachteil von CL

Da sich die Autoren an viele Regeln halten müssen, wird der Schreibprozess erheblich verlangsamt. (Die Idee einer CL besteht ja darin, Mehrdeutigkeit bereits zur Verfassungszeit zu reduzieren.) Oft müssen unzulässige (unapproved) Worte durch zulässige (approved) ersetzt werden. Derartige Ersetzungen können die Reformulierung des gesamten Satzes erfordern.

Ersetzung von unzulässigen Begriffen

1. Calibrate test set according to manufacturer's instructions.
Problem: „according to“ ist nicht zulässig (unapproved). Stattdessen soll „refer to“ verwendet werden.
2. To calibrate the test set, refer to the manufacturer's instructions.

Überprüfung der Konformität

Ein Programm unterstützt den Autor bei der Überprüfung, ob sein Text den Regeln einer CL entspricht, und unterbreitet entsprechende Korrekturvorschläge. Wenn ein Wort mehrere Bedeutungen hat, muss sich der Autor für eine entscheiden. Diesen Prozess nennt man interaktive Disambiguierung.

Die Automatisierung der Korrektur ist einfach, wenn der Ersatzbegriff derselben syntaktischen Kategorie angehört (z.B. *right* und *correct*). Schwierig sind Begriffe wie *prevent* und *preventive*, da eine Reformulierung des gesamten Satzes erforderlich ist.

CL aus Übersetzersicht

Die Verwendung von CL fördert die terminologische Konsistenz und erleichtert es dem Übersetzer / dem Autor, sich in eine unbekannte Domäne einzuarbeiten, da technische Ausdrücke ausdrücklich festgelegt werden müssen (vgl. SE). Ausserdem können Sätze für sich übersetzt werden, da relative Bezüge zwischen den Sätzen vermieden werden.

CL ist andererseits oft repetitiv, d. h. die gleichen Sätze werden immer auf dieselbe Weise übersetzt / formuliert. Aus Sicht des Übersetzers / des Autors ist das stilistisch unschön. Stilistische Variationen können aber im Widerspruch zu terminologischer Konsistenz stehen.