

ECATERINA RASCU/JOHANN HALLER

Textoptimierung durch stilistische Paraphrasen in der Technischen Dokumentation

1. Einleitung

Durch die zunehmende Technologisierung in allen Lebensbereichen wurden die Qualitätsanforderungen an produktbegleitende technische Dokumentationen erheblich erhöht. Nationale und internationale Normen (DIN EN 62079, DIN EN 12100-2), allgemein gültige Verständlichkeitsregeln sowie unternehmensspezifische Vorgaben in Redaktionshandbüchern regeln zum Teil Inhalt, Form und Sprache technischer Dokumentationen, um den sicheren Gebrauch von Produkten zu ermöglichen. Jedoch ist das Überprüfen von Texten auf normen- und firmengerechte Formulierungen eine zeitaufwendige und kostenintensive Aufgabe, und die notwendigen Umformulierungen erfordern ein hohes Maß an Übung und Erfahrung. Deshalb werden in der Praxis immer häufiger spezielle Werkzeuge wie Dokumentmanagementsysteme, terminologische Datenbanken oder Sprachkontrollwerkzeuge eingesetzt. Mithilfe dieser Werkzeuge soll sichergestellt werden, dass technische Dokumente einfacher zu lesen, zu verstehen und zu übersetzen sind.

In diesem Beitrag wird über die Weiterentwicklung des am Institut für Angewandte Informationsforschung entwickelten Sprachkontrollwerkzeugs MULTILINT berichtet, das Orthographie, terminologische Konsistenz, Grammatik und Stil eines technischen Dokumentes überprüft. Zurzeit werden fehlerhafte und stilistisch unangemessene Texteinheiten markiert und dem Autor detaillierte Fehlerbeschreibungen und typische Umformulierungsbeispiele zur Unterstützung der Korrektur zur Verfügung gestellt. Zusätzlich sollen durch die Integrierung einer Paraphrasierungskomponente automatisch Vorschläge für eine angemessenere Ausdrucksweise angeboten werden.

2. Sprachkontrollwerkzeuge für die Technische Dokumentation

Sprachkontrollwerkzeuge überprüfen einen Text nach bestimmten Kriterien und unterstützen damit den Autor bei der sprachlichen Optimierung seines

Dokumentes. Die handelsüblichen Prüfprogramme (z. B. die Rechtschreib- und Grammatikkontrolle des MS-Word-Editors, Corrigo, DUDEN-Korrektor, Deutsch Korrekt 2000) erkennen zwar Rechtschreib- und Grammatikfehler in Texten, sind aber auf den allgemeinen Sprachgebrauch zugeschnitten. Zur Unterstützung von Autoren in spezialisierteren Bereichen wie der Technischen Dokumentation wurden Werkzeuge entwickelt, die neben Rechtschreib- und Grammatikprüfungen auch Terminologie- und Stilkontrollen sowie eine Konsistenzprüfung anbieten (z. B. MAXit, SECC, KANT, MULTILINT). Idealerweise haben solche Werkzeuge zwei Komponenten, einen Checker, der Terminologie, Grammatik, Stil und Konsistenz überprüft, und eine Paraphrasierungskomponente, die Korrekturvorschläge generiert. Die Korrektur kann dann automatisch oder interaktiv erfolgen (Mitamura/Nyberg 2001).

3. Stilkontrolle mit MULTILINT

Im Folgenden werden die Stilkontrolle in MULTILINT sowie die verschiedenen Strategien, die zur Unterstützung der Korrektur benutzt werden können, kurz dargestellt.

Die Stilkontrolle in MULTILINT dient der Erkennung mehrdeutiger, unverständlicher und unangemessener Texteinheiten in technischen Dokumenten (Haller 2000). Sie basiert auf einem Regelsatz, der in Zusammenarbeit mit Spezialisten im Bereich Technische Dokumentation formuliert (Schmidt-Wigger 1998) und dann anhand von Praxiserfahrungen weiter an die Benutzerbedürfnisse angepasst wurde. Die Regeln betreffen gegenwärtig folgende Problembereiche: Layout, lexikalische und strukturelle Ambiguität, lexikalische Vagheit, strukturelle Komplexität, elliptische Konstruktionen, Wortstellung etc. Während der Stilprüfung werden unerwünschte Strukturen gesucht, mit einem Fehlercode markiert und farbig dargestellt. Der Benutzer hat dann die Möglichkeit, eine detaillierte Fehlermeldung und/oder ein typisches Umformulierungsbeispiel aus dem Fachgebiet aufzurufen, das ihn bei der Korrektur des Textes unterstützen soll (Hernandez/Rascu 2004). Des Weiteren wird gegenwärtig ein Paraphrasierungsmodul implementiert, dessen Aufgabe es ist, automatisch Umformulierungsvorschläge für die stilistisch unangemessenen Texteinheiten zu generieren.

Für die Ableitung geeigneter Paraphrasierungsregeln wurden zuerst verschiedene Korpusanalysen durchgeführt. Zum einen wurden Verstöße gegen die MULTILINT-Stilregeln in Software-Handbüchern genauer untersucht, um eine Feinkategorisierung der stilistischen Problembereiche vornehmen zu können. Zum anderen wurde ein Korpus mit typischen Umformulierungsbeispielen untersucht, wobei für die identifizierten Feinklassen typische Paraphrasierungsregeln abgeleitet wurden. Außerdem wurden bei der Formulierung der Regeln relevante Forschungsergebnisse aus den Bereichen Kontrollierte Sprachen (Lehrndorfer 1996), Texttypologie (Göpferich 2002),

Sprechakttheorie und Verständlichkeitsforschung berücksichtigt. Beispiel (1) zeigt, wie die lexikalische Vagheit bei der Benutzung des Indefinitpronomens *man* durch eine entsprechende Paraphrase aufgelöst werden kann:¹

- (1) $V_{\text{fiv,sg}} + \textit{man} \Rightarrow V_{\text{fiv,pl}} \textit{Sie}$
 # *Um sich anzumelden, gibt man das Passwort ein.*
 \Rightarrow *Um sich anzumelden, geben Sie das Passwort ein.*

Im Beispiel (2) wird durch die Verwendung des Imperativs in der Paraphrase der illokutionäre Nachdruck der ursprünglichen Äußerung präzisiert:

- (2) $NP_{\text{nom}} \textit{sein zu} V_{\text{inf}} \Rightarrow V_{\text{imp}} \textit{Sie} NP_{\text{acc}}$
 # *Die Schrauben sind zu ersetzen.*
 \Rightarrow *Ersetzen Sie die Schrauben.*

Die von der Paraphrasierungskomponente zurzeit abgedeckten stilistischen Problembereiche sind: lexikalische Vagheit, Vagheit in Bezug auf den illokutionären Nachdruck, strukturelle Komplexität, Weitschweifigkeit etc.

Die Paraphrasierungsregeln beschränken sich jedoch nicht nur auf Informationen auf der Satzebene. Auch Mehrdeutigkeiten auf der Textebene, z. B. bei anaphorischen Pronomina, werden durch entsprechende Paraphrasen aufgelöst.

Die Implementierung der Paraphrasierungsregeln beruht auf einem regelbasierten Formalismus (Carl/Schmidt-Wigger 1998), in dem Prinzipien der Mustererkennung zum Einsatz kommen. Die sequenziell ablaufenden Regeln operieren auf der Ausgabe der morphosyntaktischen Analyse eines Textes und auf den Ergebnissen der MULTILINT-Stilkontrolle. Es wird dabei nach bestimmten Mustern gesucht, deren Struktur oder Inhalt dann durch verschiedene Operationen modifiziert wird.

4. Schlussbemerkung

Bei der Anwendung des erweiterten Sprachkontrollwerkzeugs MULTILINT ergeben sich eine Reihe von Vorteilen. Abstrakte wissenschaftliche Erkenntnisse aus verschiedenen Forschungsbereichen finden ihre praktische Anwendung in den automatisch generierten Korrekturvorschlägen und werden dadurch, gemäß der Forderung von Krings (1996, S. 34), unmittelbar handlungsleitend. Außerdem können technische Autoren durch die regelmäßige Anwendung des Sprachkontrollwerkzeugs ein höheres Sprachbewusstsein entwickeln und ihre Formulierungskompetenz erweitern. Ein weiterer Effekt

¹ Für die Darstellung der Regeln wurden die folgenden Abkürzungen verwendet: P-Präposition, N-Nomen, V-Verb, acc-Akkusativ, dat-Dativ, nom-Nominativ, sg-Singular, pl-Plural, fiv-finites Verb. # markiert den stilistisch unangemessenen Satz, während \Rightarrow den automatisch generierten Korrekturvorschlag markiert.

ist die Erstellung von Texten, die stilistisch konsistent und angemessen sind. Durch Sicherung stilistischer Konsistenz sowie Reduzierung der Ambiguität und struktureller Komplexität wird auch die Arbeit der Übersetzer solcher Texte erleichtert.

Jedoch ist MULTILINT nur ein Werkzeug, das technische Autoren bei der Erstellung verständlicherer und leichter übersetzbarer Dokumentationen unterstützen soll, das aber erfahrene und sachkundige Autoren bzw. Lektoren nicht ersetzen kann (Böhler/Ceglarek 2001). Dementsprechend ist die implementierte Korrekturstrategie interaktiv: Die entdeckten Fehler und unangemessenen Strukturen werden markiert und angezeigt, und dem Autor werden Korrekturhilfen in Form von Fehlermeldungen und Paraphrasen des Ausgangstextes zur Verfügung gestellt. Dem Autor bietet sich nun die Möglichkeit, zu entscheiden, ob er den Vorschlag übernimmt oder eine andere Umformulierungsstrategie verwendet.

Literatur

- Böhler, Klaus/Ceglarek, Regine (2001): Evaluierungsbericht Lektorat. In: TETRIS Projektdokumentation 01/02. Saarbrücken: Institut für Angewandte Informationsforschung. S. 7–21.
- Carl, Michael/Schmidt-Wigger, Antje (1998): Shallow Post Morphological Processing with KURD. In: Proceedings of the Conference on New Methods in Natural Language Processing, NeMLaP'98. Sydney: ACL. S. 257–265.
- Göpferich, Susanne (2002): Textproduktion im Zeitalter der Globalisierung. Entwicklung einer Didaktik des Wissenstransfers. Tübingen: Stauffenburg.
- Haller, Johann (2000): Sprachtechnologie für die Automobilindustrie. In: Wilss, Wolfram (Hg.): Weltgesellschaft – Weltverkehrssprache – Weltkultur. Tübingen: Stauffenburg. S. 250–263.
- Hernandez, Maryline/Rascu, Ecaterina (2004): Checking and Correcting Technical Documents. In: Vienney, Séverine/Bioud, Mounira: Correction automatique: bilan et perspectives (BULAG). Besançon: Presses universitaires de Franche-Comté. S. 69–84.
- Krings, Hans P. (1996): Wieviel Wissenschaft brauchen Technische Redakteure? Zum Verhältnis von Wissenschaft und Praxis in der Technischen Dokumentation. In: Krings, Hans P. (Hg.): Wissenschaftliche Grundlagen der technischen Kommunikation. Tübingen: Gunter Narr. S. 5–128.
- Lehrndorfer, Anne (1996): Kontrolliertes Deutsch. Linguistische und sprachpsychologische Leitlinien für eine (maschinell) kontrollierte Sprache in der Technischen Dokumentation. Tübingen: Gunter Narr.
- Mitamura, Teruko/Nyberg Erik (2001): Automatic Rewriting for Controlled Language Translation. In: Proceedings of the NLP/RS 2001 Workshop on Automatic Paraphrasing: Theories and Applications. Tokyo: National Center of Sciences, S. 1–12.
- Schmidt-Wigger, Antje (1998): Grammar and Style Checking for German. In: Proceedings of CLAW '98. Pittsburgh: Language Technologies Institute, Carnegie Mellon University. S. 76–86.