

# Sprachtechnologie für die Automobildokumentation

Johann Haller  
IAI  
Martin-Luther-Str. 14  
D-66111 Saarbrücken  
hans@iai.uni-sb.de

## Kurzfassung

Die Globalisierung der Märkte hat einschneidende Veränderungen im Wirtschaftskreislauf gebracht: einerseits werden die Produkte immer differenzierter, individueller und komplizierter, andererseits muß aber auch in großen Stückzahlen rund um die Welt verkauft werden, um solche Güter in rentabler Weise zu produzieren und zu vermarkten. Während in der Herstellung die Einführung neuer Produktionsmethoden zu erheblichen Produktivitätsfortschritten geführt hat, ist die Dokumentation von technischen Produkten weiterhin ein Engpaß. Im EU-Projekt MULTIDOC (Multilinguale Dokumentation), das das Institut für Angewandte Informationswissenschaften (IAI) zusammen mit fünf europäischen Automobilherstellern durchführt, wurde dieser Bereich eingehend untersucht und gemeinsam ein prototypischer Lösungsvorschlag entwickelt. Die technische Dokumentation kann nur dann den höchsten Ansprüchen genügen, wenn sie verständlich, aktuell und mit den modernsten technischen Mitteln gefertigt, übertragen und verfügbar wird.

Dieses Ziel kann nur durch eine Integration von Dienstleistungen, Dokumentation und Netzwerklösungen und die Einbindung moderner Sprachtechnologie erreicht werden.

Im Vortrag werden einige Resultate des MULTIDOC-Projekts beschrieben und Beispiele gezeigt, wie z.B. die Autoren von Reparaturanleitungen eines großen Autoherstellers durch Sprachtechnologie bei ihrer Arbeit unterstützt werden können: eine spezialisierte Rechtschreib-, Grammatik- und Terminologiekontrolle garantiert ein fehlerfreies Dokument, das Teile und Arbeitsgänge mit den für das Unternehmen festgelegten Bezeichnungen benennt, Konsistenz- und Stilkontrolle verbessern die Lesbarkeit und Verständlichkeit des Dokuments. In der Folge werden auch erhebliche Erleichterungen und Einsparungen bei der Übersetzung erreicht; Übersetzungsgedächtnisse („Translation Memories“) erzielen eine höhere Trefferrate, die Übersetzer selbst müssen weniger rückfragen und haben für einen bestimmten Term eine einzige Übersetzung.

## Einleitung

"Übersetzung" im klassischen Sinne, d.h. die Übertragung eines fertig vorliegenden Textes in eine Fremdsprache, ist heute als Prozeß nur noch selten ein gangbarer Weg in der technischen Dokumentation. In Weltkonzernen werden Dokumentationen oft in mehreren Ursprungssprachen erstellt, und die Dokumentation soll oft gleichzeitig mit der Verkaufsbereitschaft des Produktes vorliegen. Längst hat man erkannt, daß auch die "Maschinelle Übersetzung" in ihrer bekannten Form keine Lösung für dieses Problem darstellt - auch die heute nach Jahrzehnten Entwicklung verfügbaren kommerziellen Systeme sind maximal für die rasche Inhaltsvermittlung geeignet. Nur in besonderen Fällen liefern sie

nach jahrelanger Vorbereitung Texte, die dem Übersetzer beim Verfertigen der Endversion als Vorlage Zeit sparen helfen. Die meisten Systeme, die als Übersetzungshilfen oder Übersetzungskomponenten auf dem Markt sind – z.B. PC-basierte Versionen wie Globalink, TRADOS-TM oder der IBM-Translation Manager sowie Systeme mittlerer Datentechnologie wie SYSTRAN, LOGOS, METAL (SYSTRAN und METAL - als T1 - sind seit einiger Zeit auch auf PC verfügbar) – sind mit mehr oder weniger komfortablen Schnittstellen zu den Komponenten ausgestattet, die die Dokumentenproduktion unterstützen. In keinem dieser Systeme jedoch ist die Integration solcher Übersetzungswerkzeuge und anderer Komponenten in die Benutzerumgebung befriedigend gelöst, genausowenig wie die Verknüpfung aller Werkzeuge nach den speziellen Anforderungen des Benutzers. Es gab einige wenige Versuche in dieser Richtung, so z.B. innerhalb des ESPRIT-Projekts ‘Translators Workbench’ oder innerhalb des EUROLANG-Optimizers, wo zumindest eine einheitliche Benutzeroberfläche für alle Funktionen geschaffen wurde. Rank Xerox arbeitet an Compass und TRANS als Übersetzungshilfen, die auf linguistischen Werkzeugen und linguistischer Architektur basieren. Die TRADOS ‘Translators Workbench’ integriert verschiedene Übersetzungshilfen in einer einheitlichen Benutzerumgebung.

Im EU-Projekt MULTIDOC (Multilinguale Dokumentation), das das Institut für Angewandte Informationswissenschaften (IAI) zusammen mit fünf europäischen Automobilherstellern durchführt, wurde dieser Bereich eingehend untersucht und gemeinsam ein prototypischer Lösungsvorschlag entwickelt. Dieser Vorschlag geht von der Grundhypothese aus, daß der Einsatz von Sprachtechnologie nur dann zu wesentlichen Verbesserungen führen kann, wenn er nicht an einem einzigen (späten) Punkt, der Übersetzung, erfolgt, sondern systematisch in alle Prozessschritte der multilingualen Dokumentation eingebunden ist.

In diesem Beitrag werden einige Resultate des MULTIDOC-Projekts beschrieben und Beispiele gezeigt, wie z.B. die Autoren von Reparaturanleitungen eines großen Autoherstellers durch Sprachtechnologie bei ihrer Arbeit unterstützt werden können: eine spezialisierte Rechtschreib-, Grammatik- und Terminologiekontrolle garantiert ein fehlerfreies Dokument, das Teile und Arbeitsgänge mit den für das Unternehmen festgelegten Bezeichnungen benennt, Konsistenz- und Stilkontrolle verbessern die Lesbarkeit und Verständlichkeit des Dokuments. Ebenso können auch die fremdsprachlichen Texte mit entsprechenden Tools kontrolliert werden.

## **MULTIDOC - Technische Beschreibung**

### **Benutzeroberfläche**

Das Werkzeug MULTIDOC kann als eigenständige Komponente (mit einem eigenen Editor und Symbolleiste) aufgerufen werden; externe Dateien in verschiedenen Formaten können geladen und verarbeitet werden. Ein Statusfenster zeigt an, welche Komponenten auf den vorliegenden Text schon angewendet wurden. Für die Einbindung in Dokumentenverwaltungssysteme (DVS) steht dieses Statusfenster als Ausgangspunkt zur Verfügung; die Schnittstelle z.B. zum Framemaker-basierten Redaktionssystem der Fa. BMW bildet eine von diesem exportierte Datei, die MULTIDOC zur Verfügung gestellt wird. Die Ausgabe erfolgt derzeit in eigenen MULTIDOC-Fenstern, eventuelle Korrekturen in den Originaleditoren der DVS. Technisch ist die Oberfläche als erster Prototyp in Tcl/Tk realisiert, was einen Zugriff auf die Komponenten der linguistischen Technologie ermöglicht, die auf UNIX- oder Windows-NT-Servern ablaufen. Alle Komponenten greifen auf die gleichen linguistischen Ressourcen zu, wodurch eine optimale Konsistenz gewährleistet wird.

Um optimale Flexibilität zu garantieren, sind die linguistischen Komponenten hinsichtlich der

verschiedenen Funktionen parametrisiert. Diese Parameter sind entweder vordefiniert oder können vom Benutzer eingestellt werden:

- Text-Format, z.B. SGML, HTML, rtf
- Quell- und Zielsprache, z.B. Deutsch, Englisch etc.
- Fachgebiet, z.B. Vertrieb, Entwicklung, Logistik, Marketing (hierarchisch geordnet)
- auszuführende Funktionen:
  - Prüfung von Rechtschreibung, Syntax und Stil
  - Terminologiekontrolle
  - Indexierung (Speicherung eines Ausdrucks, z.B. Abkürzungen und ihre Bedeutung, Hauptthemen des Dokuments etc.)
  - ein- oder mehrsprachiges Retrieval
  - automatische Generierung eines Glossars
  - Produktion von zusätzlichem Material für Translation Memories und Terminologische Datenbanken (Extraktion von Termkandidaten/Phrasen aus dem Text)

Für alle diese Funktionen sind die morphologische und eine flache syntaktische Analyse als Vorverarbeitung nötig. Hinzu kommt ein Formalismus, mit dem linguistische Patterns in den Analysedateien identifiziert und in bestimmter Form ausgegeben werden können.

## Linguistische Analysekomponenten

### Morphologische Analyse

Die morphologische Analyse von Texten basiert auf dem System MPRO (MAAS 96), das über eine Reihe von europäischen Sprachen linguistisches Wissen in Form von Morphemen, Allomorphen, Kombinations- und Verbotsregeln gespeichert hat. Für das Deutsche sind etwa 30.000 Morpheme notwendig, mit denen eine sehr grosse Abdeckung erreicht werden kann. Im Test mit fachspezifischen Texten müssen nur wenige neue Morpheme ergänzt werden. Im folgenden ein Beispiel für eine Analyse des Wortes 'Seitenfenster':

```
>Seitenfenster
entry = {string='Seitenfenster',lex=fenster,clex= {lex=seite,clex=nil,
hyper={f1=loc,f2=rand},head={cat=n,deriv=nil,pref=nil,ehead={gen=fem}}},
lemma=seitenfenster,head={cat=n,deriv=nil,pref=nil,ehead={gen=neuter},ehead=({cas
e=(nom;dat;acc),num=sing};{case=(nom;gen;acc),num=plu})},graphics=cap}. [] .

-- Surface structure --
[compound,n,[hyper=loc,more=way,up=an],fenster,[n,[ns,n,a,seite,(0;n)],[no,[hyper=
loc,more=rand],[n,0]],fem,no]]]

-- Semantic structure --
[np,[n,[hyper=loc,more=way,up=an],fenster],[fuehren,[n,
[hyper=loc,more=way,up=an],c0,fenster],[n,[hyper=loc,rand]wohin,seite,sg,def],no]
]

=== Interpretation: ===
Ein Seiten|fenster ist ein Fenster, das zur Seite führt.
```

Auch feste Mehrworteinheiten ('in bezug auf') sind im Morphemwörterbuch enthalten, ebenso wie Abkürzungen und Akronyme. Komplexe Komposita ('Drosselklappenzug') sind dagegen

nur mit ihren Einzelbestandteilen vertreten, da sie in ihre Einzelteile zerlegt werden können; das Gleiche gilt für regelmäßige Derivationen und ihre Kombinationen mit anderen Morphemen ('Drosselungseinheit'), da diese kompositionell analysiert werden können. Die Morpheme im Wörterbuch können auch anderen Sprachen entstammen ('Polyurethandioxidverbindung').

## Syntaktische Analyse

Die syntaktische Analyse dient dazu, jedem Wort seine eindeutige grammatische Kategorie und Funktion zuzuweisen sowie zusammengehörige Gruppen zu identifizieren, was gleichermaßen für Korrekturwerkzeuge und für Übersetzungshilfen wichtig ist. Sie besteht aus der Beschreibung grammatikalisch korrekter Strukturen sowie den Algorithmen für Analyse und Synthese. Die syntaktische Verarbeitung wurde auf der Basis ausgewählter Korpora evaluiert und vervollständigt. Besonderes Augenmerk wurde auf die Verbesserung der Effizienz gerichtet. Für das Hauptanwendungsgebiet (Reparaturanweisungen und Kundendienst-Informationen im Automobil-Sektor) wurden z.B. Formulierungen im Telegramm-Stil mitbehandelt, für den z.B. verblose Konstruktionen, subjektlose Verbalgruppen und artikellose Nominalgruppen typisch sind.

Die syntaktische Analyse läuft in mehreren Schritten ab. Einige Mehrdeutigkeiten scheiden bereits durch einfache 'Nachbarschafts'-Tabellen aus; weitere werden durch die Durchführung einer flachen syntaktischen Analyse entschieden, die Nominalgruppen und einige andere syntaktische Satzeinheiten ermittelt. Viele Analyse-Regeln können auch mit gelockerten Restriktionen arbeiten, damit z.B. später syntaktische Fehler leichter erkannt werden können. Als Beispiel für diese flache syntaktische Analyse mag der Satz dienen: 'Kundenbeschwerden sind immer Ausdruck eines subjektiv empfundenen, aktuellen oder tatsächlich vorhandenen Mangels'. Die Zusammenfassung der Gruppen wird durch entsprechende Einrückung symbolisiert.

```
<1> bis <2>:
  {c=np,r=574}
  {ori=Kundenbeschwerden,gra=cap,c=noun,lu=kundenbeschwerde,s=text,ls=kunde#beschwerde,ss=agent&
  anto#text,lngs=germ#germ,ehead={nb=plu,g=f}}
```

```
<2> bis <3>:
  {ori=sind,string=sind,lu=sein,c=w,sc=verb,vtyp=fiv,s=sein,tns=pres,mode=ind,per=3;1,nb=plu,gra
  =small}
```

```
<3> bis <4>:
  {ori=immer,string=immer,lu=immer,c=w,sc=adv,fu=pp,s=temp,gra=small}
```

```
<4> bis <5>:
  {c=np,r=574}
  {ori=Ausdruck,gra=cap,c=noun,lu=ausdruck,s=pstate,cs=n,ls=aus_&druck,ss=pstate,ehead={case=nom
  ;dat;acc,nb=sg,g=m}}
```

```
<5> bis <14>:
  {c=chapter,r=515}
    {c=np,r=203,tr=d_a_n,case=gen,nb=sg,g=m}
    {ori=eines,lu=ein,c=w,sc=art,fu=det,spec=indef,gra=small,case=gen,nb=sg,g=m}
      {c=adj,r=213,case=gen,nb=sg,g=m}
      {ori=subjektiv,gra=small,c=adv,lu=subjektiv,deg=base,s=abstract,cs=a,ls=subjekt
      ,ss=abstract}
        {c=adj,r=15}
        {ori=empfundenen,pctr=yes,gra=small,c=adj,lu=empfunden,endung=en,deg=base,ptc=2
        ,cs=v,ls=empfinden,ss=v}
          {ori=&cm,c=w,sc=comma,lu=&cm,gra=other}
          {ori=aktuellen,pctr=yes,pctl=yes,gra=small,c=adj,lu=aktuell,endung=en,deg=base,
          cs=a,ls=aktuell,ss=a}
            {ori=oder,lu=oder,c=w,sc=coord,gra=small}
            {ori=tatsalchlichen,pctl=yes,gra=small,c=adj,lu=tatsaechlich,endung=en,deg=base
            ,cs=a,ls=tatsaechlich,ss=a}
              {ori=Mangels,pctr=yes,gra=cap,c=noun,lu=mangel,s=quality,cs=n,ls=mangel,ss=quality,ca
              s=gen,nb=sg,g=m}
              {ori=.,lu=stop,c=w,sc=punct,last=yes,gra=other}
```

## Kontrollwerkzeuge und Hilfen für den Autor/Übersetzer

### Textkontrolle

Die Werkzeuge zur Textkontrolle sind optimal auf das jeweilige Fachgebiet abgestimmt und benutzen, wie oben erwähnt, einheitliche Ressourcen; die gleichen Wörterbücher und Terminologielisten werden für alle Tools von der Rechtschreibkontrolle bis zur Übersetzungshilfe verwendet. Durch diesen Ansatz und die exhaustive linguistische Analyse kann ein höherer Wirkungsgrad erreicht werden als bei der willkürlichen Kombination von kommerziellen Rechtschreib- und Grammatikhilfen in Editoren (Word, Framemaker etc.) mit Standard-Übersetzungssystemen.

So können z.B. bei der Rechtschreibprüfung genauere Verbesserungsvorschläge gemacht werden, die auf der firmenspezifischen Terminologie beruhen und nicht auf der (oft zufälligen) Ähnlichkeit von Strings.

So werden bei der (falschen) Schreibweise *Montagestellung*, die die morphologische Analyse nicht als korrekte Zusammensetzung identifizieren kann und deshalb als Rechtschreibfehler werden: das falsche Wort *Montageeinstellung* wird markiert, die terminologisch einschlägigen Möglichkeiten *Montagestellung* und *Demontageeinstellung* genannt.

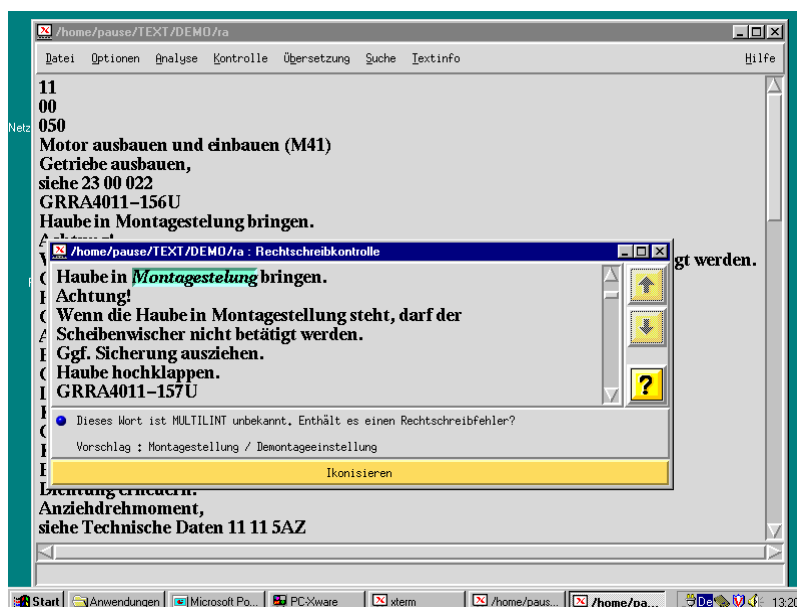


Abb.1: Rechtschreibprüfung

Die Grammatikprüfung, die sich als erstes Werkzeug des 'syntaktischen Pattern-Matchers' bedient, kann ebenfalls auf die Bedürfnisse der Textsorte zugeschnitten werden. Neben einfachen Prüfungen auf korrekte Flexion von Nominalgruppen (*der vorderer Scheibenwischer*) kann sie auch komplexere Verstöße gegen die deutsche Wortstellung entdecken (*Die Rohbautoleranzen sind eingehalten nicht worden.*).

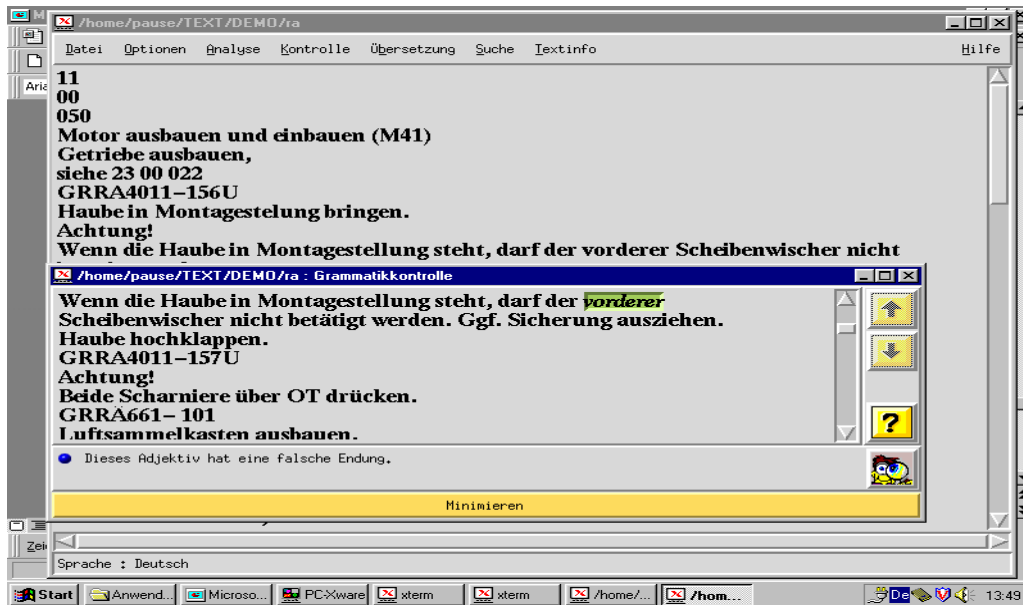


Abb. 2: Grammatikprüfung

Die Terminologieprüfung stellt zum einen die Konsistenz innerhalb von Dokumenten sicher (*Spannkolbe/Spannerkolben*); zum anderen markiert sie auch Terme, die nicht zum autorisierten Termvorrat des entsprechenden Sachgebietes gehören und sucht in diesem Termvorrat nach ähnlichen Ausdrücken, wenn nicht von vornherein eine positive Alternative festgelegt ist (statt *Luftsammlerkasten* muss es beispielsweise *Luftsammlergehäuse* heißen). Die semantische Verwandtschaft von *Kasten* und *Gehäuse* (beide sind im Wörterbuch semantisch als *container* gekennzeichnet) hilft hier zur Identifizierung des korrekten Terms.

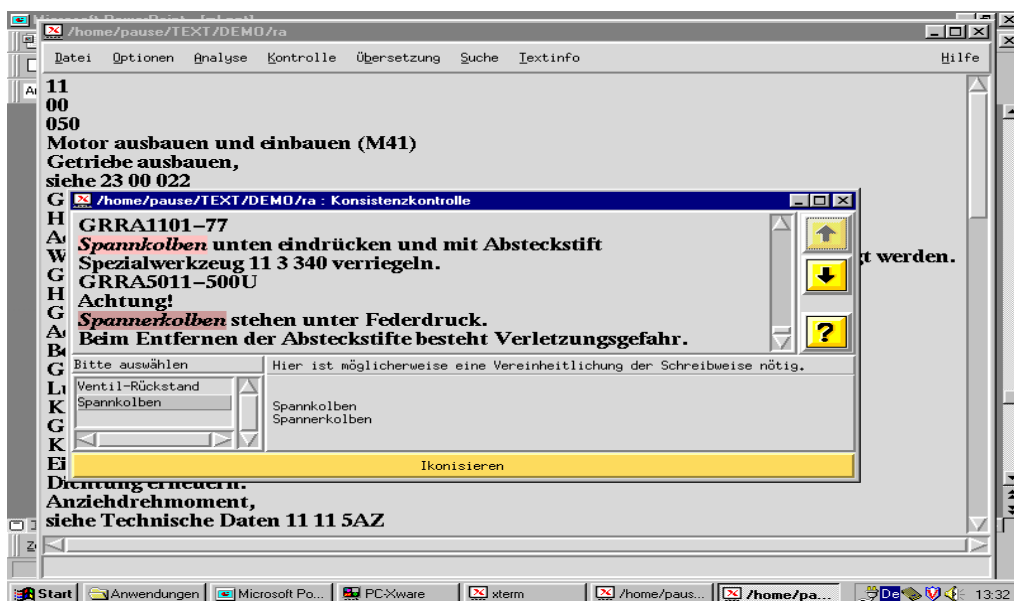


Abb.3: Konsistenzprüfung

Die Stilprüfung versucht, Verstöße gegen Regeln zu entdecken, die die Korrektheit und Verständlichkeit technischer Dokumentation garantieren sollen. Sie erkennt komplexe Wörter (*Zylinderkopfdichtfläche*) und schlägt dem Autor vor, das Wort durch einen Bindestrich in klar abgegrenzte Teile zu unterteilen. Einfach zu erkennen sind einzelne Vokabeln, die in solchen Texten nicht auftauchen sollen (Indefinitpronomen wie *man*, ungenaue Verben wie *erfolgen* etc.), während für umständliche Modalkonstruktionen wie *ist .. zu, soll + Infinitiv* wieder ein Rückgriff auf die syntaktische Struktur notwendig ist. Die Stilprüfung folgt zwar den Grundsätzen, wie sie schon vor Jahren für ein 'Kontrolliertes Englisch' entworfen und

auch in praktische Werkzeuge umgesetzt wurden (AECMA 95, Hoard 92, Wojcik 96), schreibt jedoch nicht konkrete Konstruktionen vor, sondern versucht, als falsch erkannte Muster zu markieren. Der Benutzer hat dann die Möglichkeit, ein Beispiel zur Erklärung aufzurufen, das seinen eigenen Schulungsunterlagen entnommen ist, mit denen er für das technische Schreiben ausgebildet wurde. Andere Beispiele wurden von den Autoren selbst beigesteuert (s.u.). Auch für Deutsch wurden über einen gewissen Zeitraum ähnliche Ansätze verfolgt (SDD, Schachtl 96).

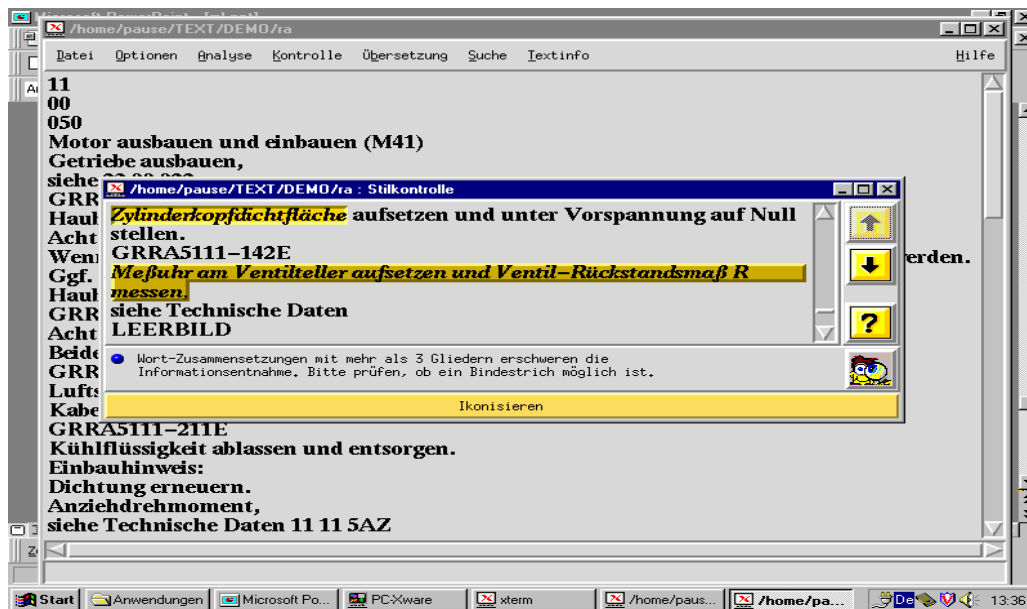


Abb.4: Stilprüfung

Die deutsche Version dieser Prüfwerkzeuge war bereits in einem vom BMWi geförderten Projekt für die Fa. BMW entwickelt worden und ist nunmehr seit einem Jahr in der gesamten Abteilung für Reparaturanleitungen und Service-Informationen (über 200 Autoren) im Einsatz. Nach anfänglicher Skepsis empfinden die Autoren MULTIDOC als eine grosse Hilfe und tragen wesentlich zur Weiterentwicklung bei. Die meisten Fehlermeldungen, die im Anfang von Entwicklern in einem theoretisch-sprachwissenschaftlichen Duktus formuliert worden waren, stammen mittlerweile von den Autoren selbst, die nach Aussage der Vorgesetzten zusätzlich ein 'linguistisches' Bewusstsein entwickelt haben und auf die sprachliche Qualität ihrer Texte mehr als vorher achten.

Neben der linguistischen Analyse ist für das Funktionieren der Prüfwerkzeuge eine wichtige Voraussetzung, dass im entsprechenden Sachgebiet möglichst vollständige Terminologielisten vorliegen. Eine in der Abteilung selbst entwickelte Terminologieliste wurde im Vergleich mit Teilverzeichnissen, mit aus Texten extrahierten Substantiven und Ausdrücken sowie besonders in Zusammenarbeit mit der Übersetzungsabteilung aktualisiert, vereinheitlicht und ergänzt. Diese Extraktion erfolgte auch aus parallelisierten Texten, wie sie in der für das Translation Memory TRANSIT verwendeten Form vorlagen. Hier wurde überprüft, inwiefern für einen deutschen Term mehrere fremdsprachliche (englische) Äquivalenzen vorlagen; wenn möglich, wurde dieser Term genauer gefasst. Wenn umgekehrt ein englischer Term die Übersetzung mehrerer deutscher Ausgangsterme war, so entschieden Kundendienst und Übersetzer zusammen, welche Variation(en) im Deutschen zum Vorzugsterm werden sollte(n).

Durch die Entwicklung englischer, französischer und schwedischer Versionen von MULTIDOC (mit Rolls-Royce, Renault und Volvo) wird es in naher Zukunft möglich sein, auch Übersetzungen der BMW-Texte in diese Sprachen zu prüfen und somit die in TRANSIT gespeicherten Übersetzungsvorschläge qualitativ auf eine höhere Stufe zu heben. Ähnliche

Projekte in umgekehrter Richtung können dann mit den ausländischen Herstellern angegangen werden.

In einer beabsichtigten Fortsetzung der MULTIDOC-Aktivitäten ist auch die Organisation der Terminologie (zunächst für ein kleines Teilgebiet) in einer sogenannten 'Ontologie' vorgesehen, einer Wissensbasis, in der die wichtigsten Entitäten und Prozesse eines Reparaturvorgangs mit ihren Abhängigkeiten dargestellt sind (Schütz 96). Wenn die Kontrollwerkzeuge auf diese Wissensbasis zurückgreifen können, wird es auch möglich sein, zu überprüfen, ob für einen bestimmten Ein- oder Ausbauvorgang die richtigen Verben den richtigen Objekten (z.B. den Kanälen - *passage* - , durch die flüssige oder gasförmige Stoffe im Auto transportiert werden) zugeordnet sind; auf der anderen Seite soll auch die Möglichkeit einer formalen Eingabe und einer anschließenden multilingualen Generierung geprüft werden. Aus existierenden Texten können die formalen Repräsentationen auch teilweise automatisch extrahiert werden.

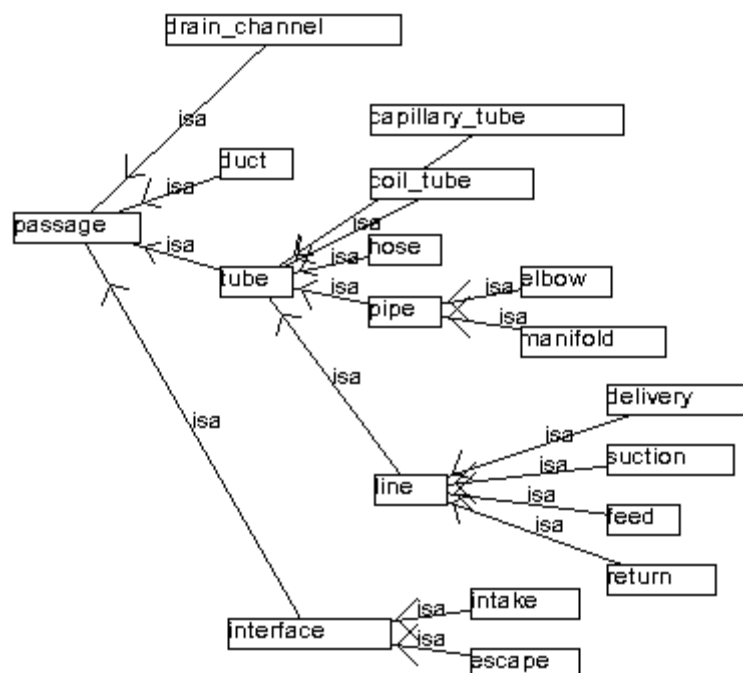


Abb. 5: Ontology passage

Bereits im laufenden Projekt werden auch die Möglichkeiten untersucht, die die Verbindung von Translation Memory und linguistischer Analyse bietet. Mit der Integration linguistischer Komponenten in solche TM-Systeme wird es möglich sein, z.B. lexikalische Elemente von morphologisch oder syntaktisch analysierten Sätzen in der Datenbank zu suchen. Wenn zu viele Sätze gefunden werden, deren Ähnlichkeit über dem prozentualen Schwellenwert liegt, kann der Benutzer eine vollständige linguistische Analyse des Textes vornehmen lassen, um Prioritäten festzulegen. Innerhalb dieser kompletten Analyse werden nicht nur lexikalische Elemente, sondern auch ihre semantischen Relationen innerhalb des Satzes auf Ähnlichkeit überprüft. Ähnliche Ziele, wenngleich auch nur auf der Basis eines statistischen Taggings, verfolgt beispielsweise das neu entwickelte Translation Memory ZERES (ZERES 1998, MOHR 1999)

### Bilinguales Glossar, Informativübersetzung und Retrieval

Aus dem Ergebnis der linguistischen Analyse kann auch ein automatisches bilinguales Glossar erstellt werden, das die lemmatisierten Wörter des Deutschen bzw. die als Terme eingetragenen Wortgruppen etwa des Englischen oder Französischen als Ausgangspunkt

nimmt. Wenn der Eintrag nicht als Ganzes in der terminologischen Datenbank enthalten ist, so werden die Äquivalente der grösstmöglichen Teile angezeigt. Ein Fachgebietsfilter kann eingeschaltet werden.

Für den Satz *Abdeckleiste aussen Seitenfenster hinten deformiert* kann z.B. das folgende Glossar erzeugt werden:

Abdeckleiste	cover strip
außen	outside
deformieren	destort
hinten	rear side
Seitenfenster	side window

Dies ist für eine Reihe von Sprachpaaren möglich, wobei für einige Zielsprachen (z.B. Russisch allerdings noch keine spezialisierte Terminologie integriert ist.

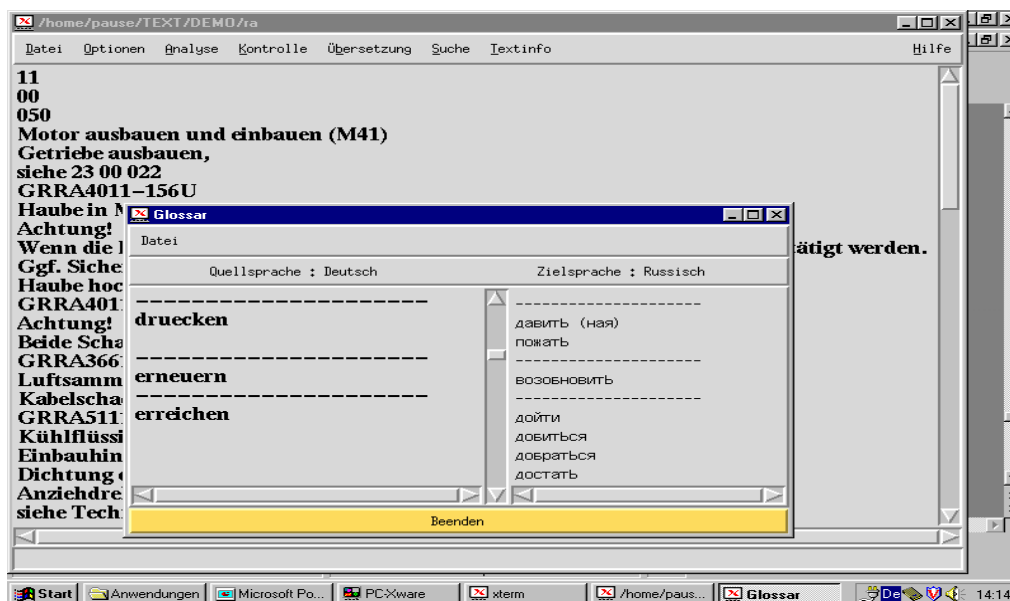


Abb. 6: Glossar Russisch

Eine andere Art der Darstellung dieser Ergebnisse ist die sogenannte 'Informativübersetzung', die die lexikalischen Entsprechungen in ihrer Textreihenfolge präsentiert und eventuell kleine Anpassungen an die Zielsprache zu machen versucht:

Cover strip outside rear side window destorted.

Diese Übersetzung dient jedoch nur zu ersten Informationszwecken bzw. der Überprüfung, ob zum entsprechenden Sachgebiet überhaupt genügend Terminologie vorhanden ist; sie wird nur in seltenen Fällen direkt als Vorlage für den Übersetzer verwendbar sein, wenn z.B. das Translation Memory keinen adäquaten Satz im Archiv gefunden hat. Hier geht MULTIDOC denselben Weg wie andere 'integrierte Systeme', wie sie z.B. von REINKE 96 beschrieben werden.

Mit Hilfe des bilingualen Glossars können auch mit dem Suchbegriff *cover strip* alle deutschen Textstellen gefunden werden, in denen *Abdeckleiste* vorkommt, und umgekehrt.

## Zusammenfassung: Nutzen für die multilinguale Textproduktion

Bei der Anwendung der beschriebenen Werkzeuge in der multilingualen Textproduktion ergeben sich nun eine Reihe von Vorteilen, die sich besonders, aber nicht nur, in der Übersetzung auswirken:

- höheres Sprachbewusstsein der technischen Autoren
- höhere Qualität der Ausgangstexte (Fehlerfreiheit, Konsistenz, Verständlichkeit)
- sprachlich einheitliches Erscheinungsbild (Image)
- höhere Trefferquote von Translation Memories (wenn Texte immer mit den gleichen Korrekturwerkzeugen überprüft werden)
- Arbeitserleichterung für den Übersetzer (weniger falsche und vage Terme und Formulierungen)
- dadurch bedingt: weniger Rückfragen nötig
- höhere Qualität der fremdsprachlichen Übersetzungsäquivalente im Archiv (durch entsprechende systematische Überprüfung der Zieltexte)

Diese Lösungen sind für die spezifischen Anforderungen eines Unternehmens speziell zu entwickeln, wobei besonders für die administrative und formale Seite (SGML) durchaus einige Standardbausteine verwendet werden können. Kommerzielle Pakete zur Sprachverarbeitung lösen dagegen heute nur Teilprobleme und erreichen deshalb meist nicht den optimalen Wirkungsgrad, weil die in ihnen enthaltenen sprachlichen Ressourcen zu breit und nicht für den Einzelfall optimiert sind. Die Ressourcen sind in diesen Werkzeugen meist so integriert, daß sie nur schwer herausgelöst und für den Benutzer angepaßt werden können. Modulare Lösungen, die auf generellen Sprachressourcen basieren, aber beim Vorliegen mehrsprachiger Texte sowie (im Idealfall) von Wörterbüchern und terminologischen Sammlungen ohne großen Aufwand für einen bestimmten Anwendungsfall angepaßt werden können, führen weitaus verlässlicher und schneller zu Rationalisierungseffekten.

Der Weg zu dauerhaften Erfolgen in der *Maschinellen Übersetzung* (wenn man den beschriebenen Weg überhaupt darunter subsumiert) kann daher, zumindest in der Industrie, nicht über die Entwicklung von Gesamtpaketen führen, die beliebige Texte aus einer Sprache blitzschnell in die andere übertragen, sondern in der modularen Integration von sprachtechnologischen Bausteinen in den gesamten Erstellungsprozess multilingualer technischer Dokumentation.

## Bibliographie

- AECMA SE-Guide (95):** 'AECMA Simplified English', AECMA Document: PSC-85-16598, AECMA, Bruxelles, 1995.
- Hoard et al. (92):** Hoard, J.E., Wojcik, R., Holzhauser, K., An Automated Grammar and Style Checker for Writers of Simplified English', in: O'Brian Holt, P., Williams, N., *Computers and Writing: State of the Art*, Kluwer Academic Publishers, Dordrecht/Boston/London, 1992, 278-296.
- Maas, D. (96):** 'MPRO-Dokumentation', in Hausser, R., *Morpholympics-Dokumentation*, Springer 1996.
- Mohr, Carsten (96):** 'Untersuchung zur Erkennung von Übersetzungseinheiten unterhalb der Satzebene in TM-Systemen'. Diplomarbeit FR 8.6. Universität des Saarlandes 06/99.
- Reinke, U. (96):** 'Integrierte Übersetzungssysteme', in: *Lebende Sprachen - Zeitschrift für fremde Sprachen in Wissenschaft und Praxis*, Nr. 3/1997.
- Schütz, J. (96a):** 'Terminology and Language Engineering', *Proceedings of TKE 96*, Vienna, August 26-28, 1996, 100-109.
- SDD-Broschüre:** 'Siemens Dokumentations Deutsch, Syntaxregeln für technische

Dokumentation', Internes Dokument, Siemens AG, München.

**Schachtl, S. (96):** 'Requirements for Controlled German in Industrial Applications', in *Proceedings of CLAW 96*, KU Leuven, Leuven, 1996, 143-149.

**Wojcik, R.H. & Holmback, H. (96):** 'Getting a Controlled Language Off the Ground at Boeing', in: *Proceedings of CLAW 96*, KU Leuven, Leuven, 1996, 22-31.

**ZERES (98):** ZeresTrans 1.2 Benutzerhandbuch, Version 20.01.1998