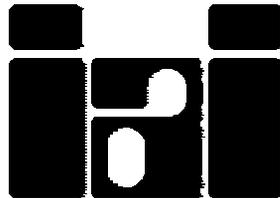




“Die Sprachwerkzeuge in int.unity”

**Abschluss Symposium
Brüssel, 28.1.2004**



Rita Nübel

IAI Saarbrücken

—

Institut der Gesellschaft zur Förderung der

Angewandten

Informationswissenschaft an der Universität des
Saarlandes

- Sprachtechnologie und multilinguales Sprach-*Engineering*
- Evaluation von Sprachtechnologie-Produkten
- Lokalisierung
- Ausbildung (UdS, Lehrstuhl für Maschinelle Übersetzung Prof. Johann Haller)

- Evaluierung Sprachtechnologie-Werkzeuge
- Ressourcen-Entwicklung
- Software-Integration MÜ-Server
- Software-Entwicklung

- Kommerzielle MÜ-Software, z.B.
 - Comprendium, Systran, Personal Translator
- Übersetzungsqualität
- Anpassungspotenzial
 - Terminologierweiterung
- Integrationsfähigkeit mit BSCW für verteilte Übersetzungsaufgaben

Korpuserstellung relevanter Texte als Basis für int.unity-Terminologie.

- Semi-automatische Extraktion von Termkandidaten.
- Semi-automatische Übersetzung durch Abgleich mit vorhandenen Ressourcen.
- Kodierung der zweisprachigen Terminologie im MÜ-Format zur Abdeckung der int.unity-Themenbereiche der Gewerkschafts-Community.

- Installation MÜ-Server am IAI
- Implementierung der Client-Server-Kommunikation
- Programmierung der Übersetzungsfunktionen, die auf Anfragen der BSCW-Plattform reagieren:
 - „TransText“ (übersetzt Notizen und Freitext)
 - „TransDoc“ (übersetzt html-Dokumente)

- **Software-Entwicklung:** Einsatz von IAI Sprachkontroll-Software CLAT.

- **Software-Entwicklung:** Einsatz von IAI Sprachkontroll-Software CLAT.
- Validierung von Ideen und Konzepten der *Kontrollierten Sprache*
 - Eindeutigkeit
 - Verständlichkeit
 - Übersetzbarkeit
- Anpassungspotenzial für nicht-technische Texte.
- Anpassungspotenzial für die Nachfolgeanwendung Maschinelle Übersetzung.

- Werkzeug für Technische Redakteure.
- Unterstützung bei der Erstellung qualitativ hochwertiger Dokumentation.
- Qualität durch
 - sprachliche Korrektheit (Rechtschreibung, Grammatik)
 - firmenspezifische Vorgaben für Terminologie und Schreibstil.
- Automatische Überprüfung von Dokumenten.
- Hinweise für den Autor
 - Fehlermeldungen
 - Korrekturbeispiele.



Projektziel aus Sicht der Sprach-Technik:

“Mehrsprachige Kommunikation und Kooperation geografisch verteilter Teams ermöglichen und optimieren”

Zwei ST-Prozesse im int.unity-Szenario:

- Software-gestützte Sprach*kontrolle* mit CLAT
 - Vorbereitung von Texten für die Maschinelle Übersetzung
- Vollautomatische Übersetzung per Knopfdruck

Ziel:

- Unterstützung bei der Erstellung von *MÜ-freundlichen* Texten
- Produktion verständlicher Übersetzungen

- **Maschinelle Sprachkontrolle mit CLAT:**
 - Rechtschreibung
 - Grammatik
 - Terminologie
 - Stil

- **Maschinelle Übersetzung mit dem Personal Translator**
 - ... *“übersetzt alles für das Sprachpaar Deutsch \Leftrightarrow Englisch”*

■ Annahmen:

“Eine sprachlich korrekte Aussage ist besser *verständlich* als eine Aussage mit Rechtschreib- und Grammatikfehlern.”

“Eine klar formulierte Aussage ist besser *verständlich* als eine umständlich, kompliziert und wortreich formulierte Aussage.”

“Eine *verständliche* Aussage kann besser übersetzt werden als eine weniger verständliche Aussage.”

- Wörter mit Rechtschreibfehlern oder Tippfehlern werden von MÜ generell nicht erkannt und daher nicht übersetzt.
- Viele Wörter - vor allem Fachbegriffe - müssen dem MÜ-System zunächst „bekannt gemacht“ werden, sonst werden sie falsch oder gar nicht übersetzt.
- Sätze mit Grammatikfehlern werden von MÜ häufig falsch analysiert und daher falsch übersetzt.
- Komplizierte Sätze werden von MÜ meistens falsch analysiert und daher falsch übersetzt.

Maschinelle Übersetzbarkeit von Texten?

- Rechtschreibung?
- Grammatik?
- Satzbau?
- Satzlänge?
- Eindeutige Formulierungen?

Maschinelle Übersetzbarkeit von Texten!

- Rechtschreibfehler vermeiden
- Grammatikfehler vermeiden
- Einfacher Satzbau
- Kurze Sätze
- Eindeutige Formulierungen

Einige Beispiele: Rechtschreibfehler

- **Eingabe:** „Wir brauchen einen Betriebrat!“
- **MÜ:** "We need a Betriebrat!"

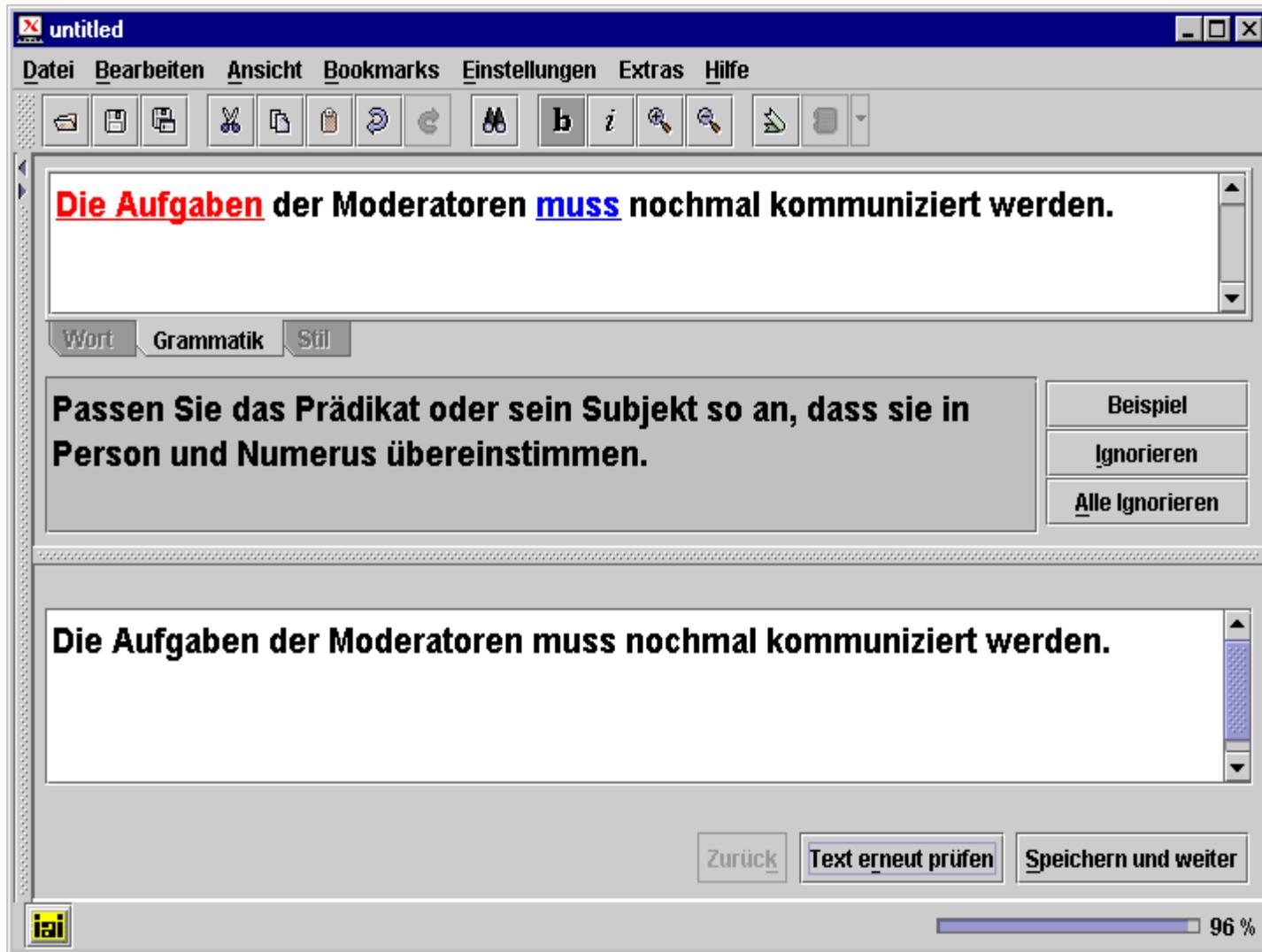
- **Eingabe:** "Trade unios and the new economy."
- **MÜ:** "Handele unios und die neue Wirtschaft."

- **Eingabe:** „Wir brauchen einen Betriebsrat!“
- **MÜ:** „We need a works council!“

- **Eingabe:** “Trade unions and the new economy.“
- **MÜ:** “Gewerkschaften und die neue Wirtschaft.“

- **Eingabe:** Die Aufgaben der Moderatoren muss nochmal kommuniziert werden.
- **MÜ:** The tasks of the moderators needs communicated be once again.

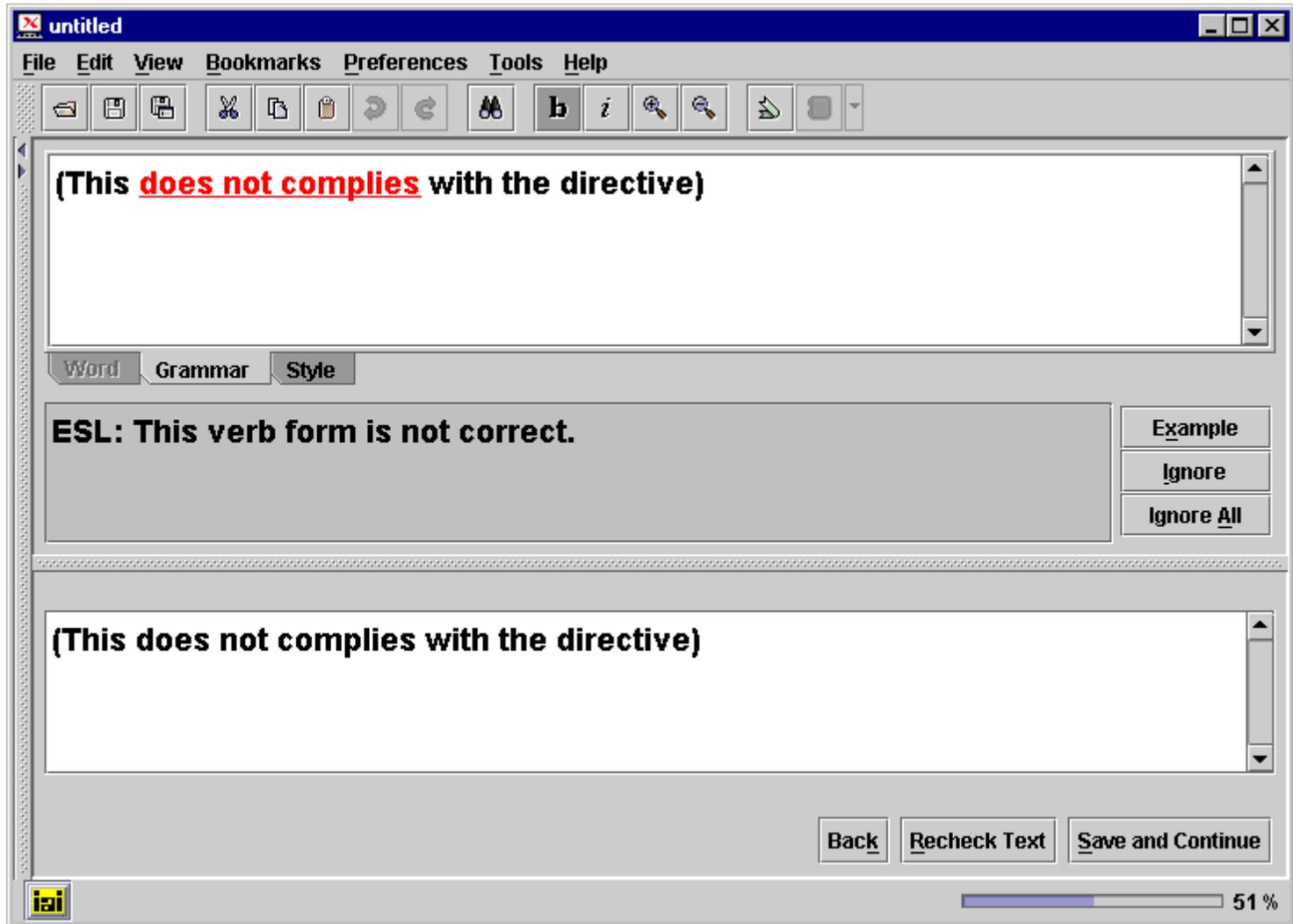
... Grammatikfehler mit CLAT erkannt ...



- **Eingabe:** Die Aufgaben der Moderatoren müssen nochmal kommuniziert werden.
- **MÜ:** The tasks of the moderators must be communicated once again.

- **Eingabe:** This does not complies with the directive.
- **MÜ:** Dies tut es nicht entspricht der Richtlinie.

... Grammatikfehler mit CLAT erkannt...



- **Eingabe:** This does not comply with the directive.
- **MÜ:** Dies entspricht der Richtlinie nicht.

- **Eingabe:** Das Projekt E-Union ist eine Vision, von der wir noch weit entfernt sind.
- **MÜ:** The project E-Union is a vision, the we still far away are.

... Stilfehler mit CLAT erkannt ...

untitled

Datei Bearbeiten Ansicht Bookmarks Einstellungen Extras Hilfe

Das Projekt E-Union ist eine **Vision, von der** wir noch weit entfernt sind.

Wort Grammatik **Stil**

Vermeiden Sie diese komplexe Struktur Satz oder teilen Sie den Satzinhalt auf auf. Danach nochmal mit CLAT prüfen.

Das Projekt E-Union ist eine Vision, von

untitled

Beschreibung:

Fehlerbeispiel:

Die Gewerkschaften werden auch zukünftig die humanen Interessen der **Menschen, für die Arbeit** die Grundlage ihrer Lebensplanung, vertreten.

Korrekturbeispiel:

Die Gewerkschaften werden auch zukünftig die humanen Interessen der Menschen vertreten, **deren Arbeit die Grundlage ihrer Lebensplanung ist.**

OK

Eingabe: Die E-Union ist eine Vision, die noch weit entfernt ist.

MÜ: The project E-Union is a vision which is still far away.

- **Eingabe:** I'm happy to provide examples, more information or clarification on any of the areas included in this report, and hope it gives an idea of our work in developing the e-union.

- ? ■ **MÜ:** Ich freue mich, Beispiele, weitere Information oder Klärung auf einigen der in diesen Bericht einbezogenen Bereiche bereitzustellen und zu hoffen, dass sie eine Vorstellung von unserer Arbeit in Entwicklung gibt, das elektronische Gewerkschaft.

... Beispiel Stilfehler Englisch mit CLAT erkannt ...\1

untitled

File Edit View Bookmarks Preferences Tools Help

I'm happy to provide examples, more information or clarification on any of the areas included in this report and hope it gives an idea of our work in developing the e-union.

Word Grammar Style

Sentences should not be longer than 20 words. Rewrite the sentence or break the sentence into several shorter sentences. Then check it / them again.

Example
Ignore
Ignore All

I'm happy to provide examples, more information or clarification on any of the areas included in this report and hope it gives an idea of our work in developing the e-union.

Back Recheck Text Save and Continue

96 %

... Beispiel Stilfehler Englisch mit CLAT erkannt ...\2

The screenshot shows a text editor window titled 'untitled' with a menu bar (File, Edit, View, Bookmarks, Preferences, Tools, Help) and a toolbar. The main text area contains the sentence: "I'm happy to provide examples, more information or clarification on any of the areas included in this **report and hope it** gives an idea of our work in developing the e-union." The words "report and hope it" are highlighted in red. Below the text area are tabs for "Word", "Grammar", and "Style". A dialog box is open over the "Grammar" tab, showing a description of the error, an example error, and an example rewrite.

untitled

File Edit View Bookmarks Preferences Tools Help

I'm happy to provide examples, more information or clarification on any of the areas included in this **report and hope it** gives an idea of our work in developing the e-union.

Word Grammar Style

Ensure that words that are linked together by "and" same part of speech, for example, noun or verb. Correct the sentence. Then check it again.

I'm happy to provide examples, more information or clarification on any of the areas included in this **repo** idea of our work in developing the e-union.

untitled

Description:

Example error:

Splitting and merging is done through the **menu and allows** you to give file systems a separate mount point.

Example rewrite:

Splitting and merging is done through the menu. This process allows you to give file systems a separate mount point.

OK

... Beispiel Stilfehler Englisch mit CLAT erkannt ...\3

The screenshot shows a text editor window titled 'untitled'. The main text area contains the sentence: "I'm happy to provide examples, more information or clarification on any of the areas included in this [report](#) and **hope it** gives an idea of our work in developing the e-union." The word 'it' is highlighted in red. Below the text area, there are tabs for 'Word', 'Grammar', and 'Style'. The 'Grammar' tab is active, showing a message: "Use of 'it' should be restricted, especially at the beginning of a sentence." Below this message, the original sentence is repeated. At the bottom right of the editor, there are 'Back' and 'Re' buttons. In the bottom left corner, there is a small 'iai' logo.

The dialog box, titled 'untitled', provides details for the error. It contains the following sections:

- Description:** Avoid the personal pronoun, possessive pronoun or demonstrative pronoun, unless the noun to which the pronoun refers is clear. Typical personal pronouns with an indefinite
- Example error:** **It** also describes the ways to install Solaris software.
- Example rewrite:** **This chapter** also describes the ways to install Solaris software.

An 'OK' button is located at the bottom right of the dialog box.

- **Eingabe:** I'm happy to provide examples, more information or clarification on any of the areas included in this report. I hope the report gives an idea of our work in developing the e-union.
- **MÜ:** Ich freue mich, Beispiele, weitere Information oder Klärung auf einigen der in diesen Bericht einbezogenen Bereiche bereitzustellen. Ich hoffe, dass der Bericht eine Vorstellung von unserer Arbeit beim Entwickeln der elektronischen Gewerkschaft gibt.

- Mehrsprachige gewerkschaftliche Kooperation wird in int.unity ermöglicht und unterstützt durch Einsatz und Integration von Werkzeugen:
- Kooperationsplattform BSCW
- Sprachwerkzeuge CLAT und MÜ

... zum Thema Sprachtechnologie

<http://www.iai-sb.de>