

“The Language Tools in int.unity”

**Final Symposium
Brussels, 28.1.2004**



Rita Nübel

IAI Saarbrücken

—

Institute of the Society for Promoting

Appplied

Information Sciences at the University of the
Saarland

- Language technology and multilingual language *engineering*
- Evaluation of language technology products
- Localization
- Training (UdS, Chair for Automated Translation Prof. Johann Haller)

- Evaluation of language technology tools
- Resource development
- Software integration in AT server
- Software development

- Commercial AT software, e.g.
 - Comprendium, Systran, Personal Translator
- Translation quality
- Adaptability
 - Terminology expansion
- Ability to integrate with BSCW for distributed translation tasks

Resources Development



Pioneer work

- Creation of a body of relevant texts as the basis for int.unity terminology.
- Semi-automatic extraction of term candidates.
 - Semi-automatic translation via comparison with existing resources.
 - Encoding of the bilingual terminology in AT format to cover the int.unity spectrum of topics in the trade union community.

- Installation of AT server at IAI
- Implementation of client-server communication
- Programming the translation functions that react to queries on the BSCW platform:
 - “TransText” (translates notes and free text)
 - “TransDoc” (translates html documents)

Software Development

Controlled
Language
Authoring
Technology



- **Software development:** use and extend IAI language control software CLAT.

- **Software development:** use and adaptation of IAI language control software CLAT.
- Validation of ideas and concepts of *controlled language*
 - Clarity
 - Comprehensibility
 - Translatability
- Adaptation potential for non-technical texts.
- Adaptation potential for the subsequent application of automated translation.

CLAT in a Nutshell

- Tool for technical writers.
- Support for writing high quality documentation.
- Quality through
 - Linguistic correctness (spelling, grammar)
 - Company-specific style requirements for terminology and writing style.
- Automatic document checking.
- Help for the author
 - Error reports
 - Correction examples.

Project aim from the language technology viewpoint:

“To enable and optimize multilingual communication and cooperation between geographically distributed teams”

Two LT processes in the int.unity scenario:

- Software-supported language *control* with CLAT
 - Preparation of texts for automated translation
- Fully automatic translation at the touch of a button

Aim:

- Support in the writing of *AT-friendly* texts
- Production of comprehensible translations

- Automated language control with **CLAT**:
 - spelling
 - grammar
 - terminology
 - style

- Automated translation with the **Personal Translator**
 - ... *“translates everything German \Leftrightarrow English”*

- **Assumptions:**

“A linguistically correct statement is more *comprehensible* than a statement containing spelling and grammatical errors.”

“A clearly worded statement is more *comprehensible* than an awkward, complicated and wordy statement.”

“A *comprehensible* statement can be better translated than a less comprehensible statement.”

“Comprehensibility” and AT?

- Words containing spelling errors or typos are generally not recognized by AT and therefore not translated.
- Many words – especially technical terms – have to first be “introduced” to the AT system, otherwise they are translated incorrectly or not at all.
- Sentences with grammatical errors are often falsely analyzed by the AT and hence incorrectly translated.
- Complicated sentences are usually falsely analyzed by the AT and hence incorrectly translated.

Automated Translatability of Texts?

- Spelling?
- Grammar?
- Sentence structure?
- Sentence length?
- Clear wording?

Automated Translatability of Texts!

- Avoid spelling errors
- Avoid grammatical errors
- Simple sentence structure
- Short sentences
- Clear wording

Some Examples: Spelling Errors

- **Entry:** “Wir brauchen einen Betriebrat!”
- **AT:** “We need a Betriebsrat!”

- **Entry:** “Trade unios and the new economy.”
- **AT:** “Handele unios und die neue Wirtschaft.”

... Spelling Errors Corrected

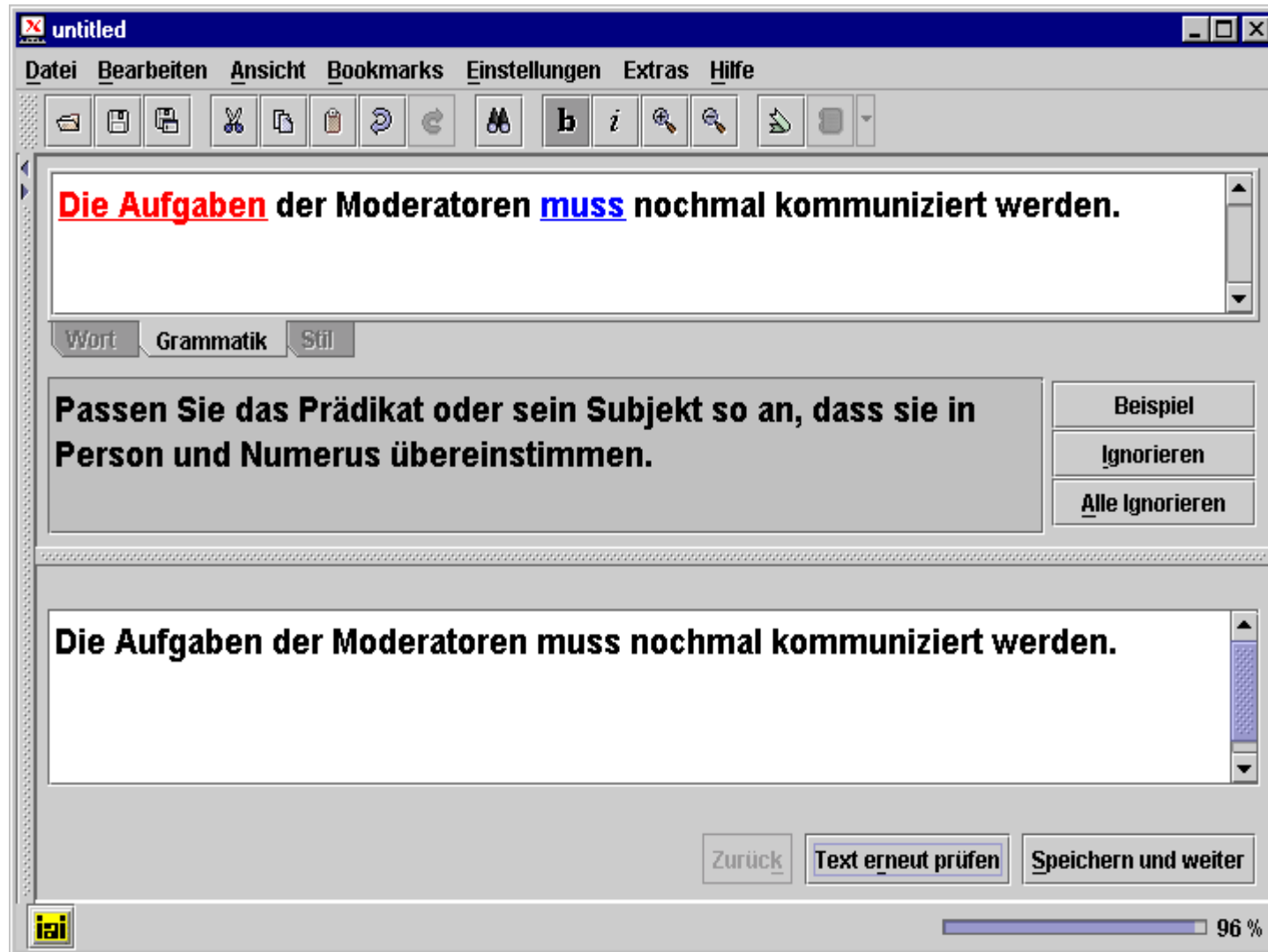
- **Entry:** “Wir brauchen einen Betriebsrat!”
- **AT:** “We need a works council!”

- **Entry:** “Trade unions and the new economy.”
- **AT:** “Gewerkschaften und die neue Wirtschaft.”

Example: Grammatical Errors

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

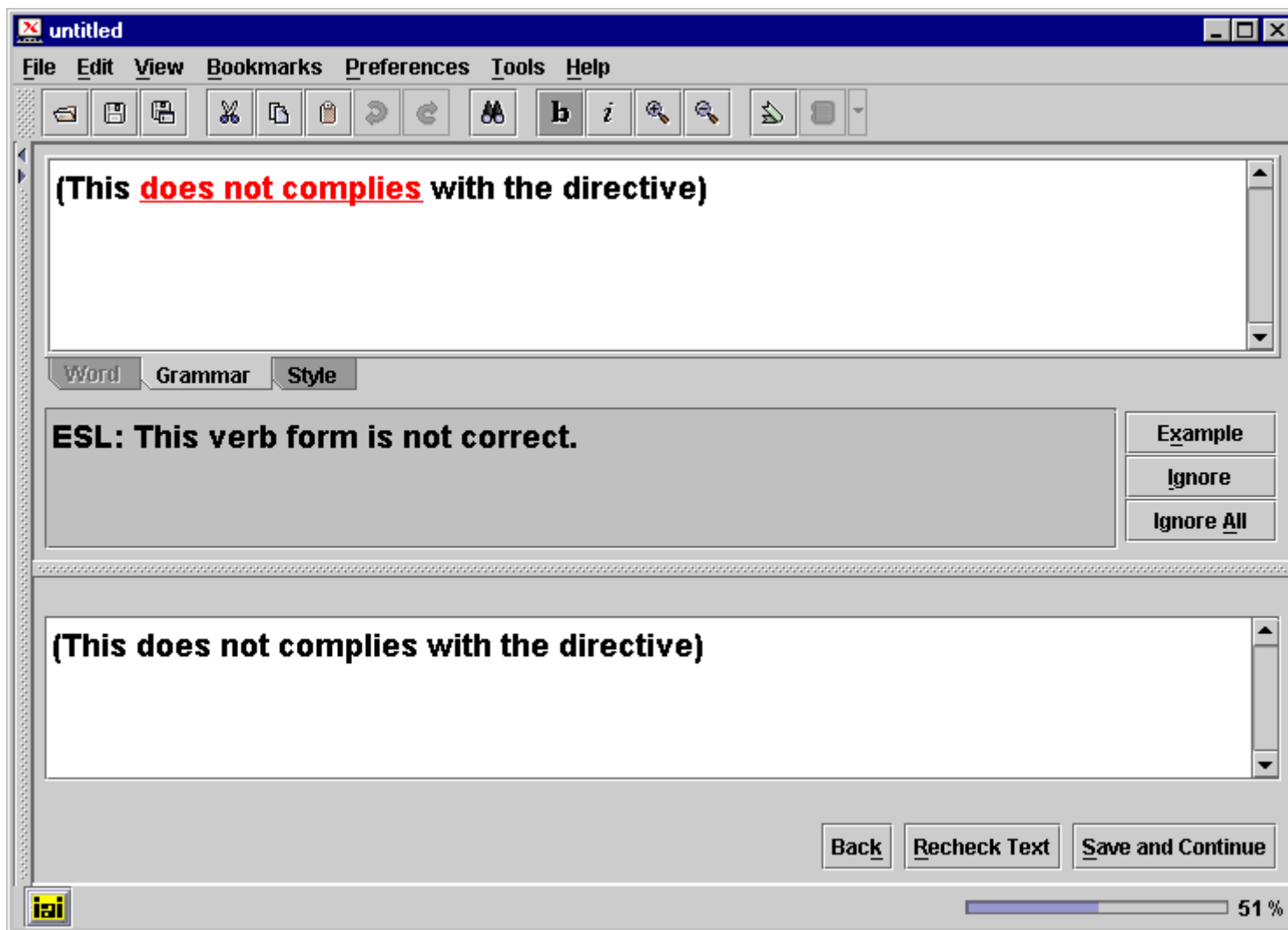
... Grammatical Errors Recognized by CLAT ...



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

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

... Grammatical Error Recognized by CLAT...



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

Example: Style Error ...

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

... Style Error Recognized by CLAT ...

The screenshot shows a text editor window titled 'untitled' with a menu bar (Datei, Bearbeiten, Ansicht, Bookmarks, Einstellungen, Extras, Hilfe) and a toolbar. The main text area contains the sentence: 'Das Projekt E-Union ist eine **Vision, von der** wir noch weit entfernt sind.' The words 'Vision, von der' are highlighted in red. Below the text area are tabs for 'Wort', 'Grammatik', and 'Stil'. A dialog box is open over the 'Stil' tab, showing a description of the error, a 'Fehlerbeispiel' (error example) with the original sentence, and a 'Korrekturbeispiel' (correction example) with the corrected sentence: 'Das Projekt E-Union ist eine Vision, von der wir noch weit entfernt sind. Die Gewerkschaften werden auch zukünftig die humanen Interessen der Menschen, für die Arbeit die Grundlage ihrer Lebensplanung, vertreten. Die Gewerkschaften werden auch zukünftig die humanen Interessen der Menschen vertreten, deren Arbeit die Grundlage ihrer Lebensplanung ist.'

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

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

- **Entry:** 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.
- ? ■ **AT:** 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.

... Example Style Error English Recognized by CLAT ...\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 %

... Example Style Error English Recognized by CLAT ...\2

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 words "report and hope it" are highlighted in red. Below the text area, there are tabs for "Word", "Grammar", and "Style". The "Style" tab is active, showing a message: "Ensure that words that are linked together by 'and' same part of speech, for example, noun or verb. Cor the sentence. Then check it again." Below this message, the same sentence is shown again, but with "report and hope it" split into "report" and "allows", indicating the correction.

The dialog box, titled 'untitled', provides details for the style error. It has three sections: "Description:" (empty), "Example error:" (containing the text "Splitting and merging is done through the **menu and allows** you to give file systems a separate mount point." with "menu and allows" in red), and "Example rewrite:" (containing the text "Splitting and merging is done through the menu. This process allows you to give file systems a separate mount point."). An "OK" button is at the bottom right.

- **Entry:** 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.

- **AT:** 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.

Summary

- Multilingual trade union cooperation is enabled and supported in int.unity through the use and integration of the tools:
- BSCW cooperation platform
- CLAT and AT language tools

Further Information

... on language technology

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