WHY TRANSLATOR-FRIENDLY?

ReEscreve is intended to:
- Help translators to fine-tune the source text to their special translation needs
- Help make the source text more consistent, more fluent and in compliance with the quality control requirements of the target language
- Help speed up and facilitate translation tasks by allowing the translator to choose the expressions that are more appropriate for the text being translated
- Help provide alternatives and probably more useful meanings in the context of a professional document (instead of the most obvious or literal meanings)
- Fill the gap for translators who are not true linguists, but terminology specialists in specific fields
- Help review the target text – can be used in automated post-editing
- Be combined with and supplement translation memories and MT
- Be used in automated post-editing specifically in specific fields
- Help translators in the authoring process
- Help translators in the post-editing process
- Help translators in the style editing process
- Help translators in the automated generation of style editors
- Help translators in the automated control of style

WHY MULTI-PURPOSE?

ReEscreve can be used for various different purposes:
- Authoring-aid
- Stylistic editing
- Pre and post-editing for MT
- Controlled language
- Automated style editors
- Evaluation of the language control process
- Importance of quality control
- MT multi-purpose
- Why translator friendliness?
- Why multi-purpose?
- Functions and relevance to NLP and (machine) translation
- Example – word re-edition
- Etc.

OUTLINE

A TRANSLATOR-FRIENDLY MULTI-PURPOSE PARAPHRASING SOFTWARE TOOL
**Current Characteristics**

ReEscreve is a prototype – with an interface – paraphraser that provides suggestions and helps users to change, simplify or clarify their texts. It has only been tested with support verb constructions (SVC), but it is being extended to progressively paraphrase different types of linguistic phenomena (adverbials, passives, relatives, etc.). At the current stage, it transforms, with a 93.4% precision, SVC into verbs or semantically equivalent expressions to help reduce ambiguity and the number of words, and eliminate unnecessary complexity. It works bilingually (EN-PT) for SVC – additional testing and improvements are in progress. It still requires linguistic quality assurance – it runs on lexica that need to be extended and enriched. Expressions must be extended to support verb constructions (SVC), which is part of the current effort. It helps ease ambiguity and the number of words and eliminates unnecessary complexity.

**Bilingually**

ReEscreve is freely accessible on the Internet at: [http://www.linguateca.pt/ReEscreve/](http://www.linguateca.pt/ReEscreve/). ReEscreve is proposed in Barreiro, 2008 as a tool for MT and MT systems. It has been experimentally applied to a prototype MT system and is being tested for quality control. Linguistic quality control is important for linguistic phenomena, such as concordance, and application to controlled writing and post-editing. However, in the future, there is a need for proper tooling that provides suggestions and helps users to prioritize clarity and integrity of language.
Current tools contemplate spell checking, simple grammar checking, orthographic and morphological variance, and synonyms. Future tools will require the following techniques:

- Word reduction = simplification
- Linguistic information for a better understanding of the meaning of the sentence, such as lexical realization of null subject in certain cases
- Word order (passive into active, etc.)
- Word and expression replacement with synonyms and paraphrases
- Alternative expressions (≠ syntax) to express the same meaning (= semantics)

Linguistic richness = many ways of "saying the same thing" using different synonyms and morphological variance, and synonyms. Current tools complement spell checking, simple grammar checking, and synonyms.

**AUTOMATED STYLE EDITORS**

**PARAPHRASING FUNCTIONS**
OUTLINE

- ReEscreve: Resources and Interface
  - Examples – word reduction
  - Paraphrases
  - Functions and relevance to NLP and (machine) translation
- Importance of Linguistic quality control
- Motivation for its development
- Current characteristics and the future?

PORT4NOOJ RESOURCES

ReEscreve: A Tool Built on Top of the

Examples of Paraphrases – Word Reduction

ReEscreve

PORT4NOOJ large coverage resources developed by

ReEscreve's paraphrasing resources were built on

The Portuguese people

The British queen

Our difficulties

The role of Europe

The speed of light

PARAPHRASE

Linguistic

Examples

Noun Phrase

Entity

Named

Clause

Relative

Adverbial

Linguistic

Expression or sentence

PARAPHRASE

Interlinear

Expression or sentence
PORT4NOOJ:
RESOURCES FOR PORTUGUESE PROCESSING

MT Summit XII
Anabela Barreiro & Luís Miguel Cabral
Ottawa, August 29, 2009

• a general dictionary of approximately 40,000 single word entries which includes:
  – (optional) inflectional paradigm information
  – (optional) derivational paradigm information
  – syntactic-semantic information
  – (optional) support verb information
  – (optional) argument-structure information
  – English translation for each entry

• 13,000 compounds (need to be assigned correct inflectional paradigms)

• morphological rules which cover inflection and derivation

• a grammar to process contractions

• syntactic grammars to recognize, annotate, paraphrase and translate support verb constructions, adverbs, relatives, passives, and possessives

• a grammar to annotate and disambiguate pronouns

• a grammar to recognize and annotate named entities

• a grammar to recognize and translate dates from Portuguese into English

DATA ON PORT4NOOJ

VALIDATION DISCLAIMER
• These resources still require extensive validation and evaluation
• We do not know as yet if we will get funding for it

ENHANCED DICTIONARY ENTRIES WITH SYNTACTIC-SEMANTIC INFORMATION

adaptar, V+FLX=FALAR+Aux=1+INOP57+Subset132+EN=

amputar, V+FLX=FALAR+Aux=1+OBJTRundif21+BioMed+EN=amputate+

amputar, V+FLX=FALAR+Aux=1+OBJTRundif21+BioMed+EN=amputate+

rápido, A+FLX=RÁPIDO+PV+eagerType+EN=quick+

adoçar, V+FLX=COMEÇAR+Aux=1+OBJTRundif75+Subset604+EN=sweeten+

transplantar, V+FLX=FALAR+Aux=1+RECTR26+Subset=504+BioMed+EN=transplant+

médico, N+FLX=ANO+AN+des+med+AG+EN=doctor

doente, N+FLX=ANO+AN+des+med+PAT+EN=patient

PARAPHRASE SUPPORT VERB CONSTRUCTIONS

GRAMMAR AND CONCORDANCE TO RECOGNIZE AND PARAPHRASE SUPPORT VERB CONSTRUCTIONS

Recognition and paraphrasing of support verb constructions (SVC/ morphossyntactic and semantically related verbs)

Controlled Language & Text Editing

We do not know if we will get funding for it

The resources feature extensive validation and evaluation

A general dictionary of approximately 40,000 single word entries which includes:

– (optional) inflectional paradigm information
– (optional) derivational paradigm information
– syntactic-semantic information
– (optional) support verb information
– (optional) argument-structure information
– English translation for each entry

13,000 compounds (need to be assigned correct inflectional paradigms)

Morphological rules which cover inflection and derivation

A grammar to process contractions

Syntactic grammars to recognize, annotate, paraphrase and translate support verb constructions, adverbs, relatives, passives, and possessives

A grammar to annotate and disambiguate pronouns

A grammar to recognize and annotate named entities

A grammar to recognize and translate dates from Portuguese into English
PARAPHRASING FOR LINGUISTIC QUALITY CONTROL AND TEXT STYLISTICS

GRAMMAR TO RECOGNIZE AND PARAPHRASE BIOMEDICAL SVC

ARGO = AGENT (AG)
Elementary SVC > Lexical Verb - fazer uma amputação = amputar (to amputate)
Elementary SVC > non-elementary SVC - realizar/efectuar uma amputação (to perform an amputation)

ARGO = PATIENT (PAT)
Submeter-se/ser submetido a uma operação = submeter-se a uma operação (to undergo surgery)
Ser operado = operado (to be operated)
Elementary SVC > non-elementary SVC - realizar/efectuar uma operação (to perform an operation)

ReEscreve architecture

ReEscreve online

Noojapply

Port4Nooj

Dictionaries

Grammars

Input Text

Suggestions

Paraphrase Suggestions
Let's try this text!

ReEscreve – INTERFACE

Interactive use of ReEscreve

ReEscreve can be used in word processing applications or linguistic quality control for both source and target texts.

Rewriting your own suggestions

User can suggest new expressions!

Log

• Original sentence: Dizia que as escadas eram uma maçada, que o senhorio estava a fazer planos para instalar um elevador, o que seria uma vantagem quando Frances fosse lá ficar com o bebé.

• Selected: Dizia que as escadas eram uma maçada, que o senhorio estava a planear instalar um elevador, o que seria uma vantagem quando Frances fosse lá ficar com o bebé.

Registering user choices and improving the resource

Original sentence: Ultimamente comecei a dar ainda mais valor à privacidade e ao anonimato desta casa.

Suggestion: Ultimamente comecei a apreciar à privacidade e ao anonimato desta casa.

MT Summit XII
Anabela Barreiro & Luís Miguel Cabral
Ottawa, August 29, 2009
Support verbs recognition of SVC

Recall
Paraphrasing of SVC

- 73/73 - 100%
- 75/75 - 100%
- 65/65 - 100%
- 57/60 - 95%
- 43/45 - 95.5%

Evaluation of recognition and paraphrasing of SVC with ReEscreve 500 sentences - manually annotated 100 sentences for each support verb

PRELIMINARY EVALUATION
Enlargement and improvement of Port 4 NooJ’s semantic relations

• Assignment of thematic roles and semantic relations to named entities
  Exs:
  médico + AG + EN = doctor
  doente + PAT + EN = patient

• Enlargement of the syntactic-semantic relations between predicates and other elements semantically related to them
  Exs:
  operar – operação – operado
  urgente – urgentemente

• Inserting synonyms, style variants, aspectual variants and enlargement of the links between verbs and support verb constructions

Many linguistic phenomena need to be formalized – one phenomenon at a time, makes the linguistic knowledge database to grow!

HOW TO IMPROVE REESCREVE?

HOW TO INTEGRATE REESCREVE IN OTHER TOOLS?

RELEVANT PUBLICATIONS


PROJECT REESCREVE is partially supported by Linguateca, jointly funded by the Portuguese Government, the European Union (FEDER and FSE), under contract ref. POSC/339/1.3/C/NAC, and by UMIC and FCCN. The first author was supported by PhD grant SFRH/BD/14076/2003.

ACKNOWLEDGEMENTS
ATRANSLATOR-FRIENDLYMULTI-PURPOSE PARAPHRASING SOFTWARE TOOL

A TRANSLATOR-FRIENDLY MULTI-PURPOSE REESCREVE