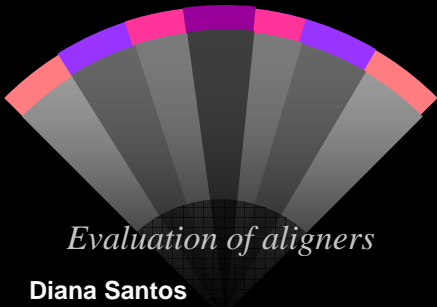


2006-10-10



Evaluation of aligners

Diana Santos
SINTEF Telecom and Informatics

2006-10-10

Presentation plan

Evaluation: Motivation	Context: CPP project and COMPARA / DISPARA
Alignment questions	Evaluation proper

2006-10-10

Evaluation

- ♦ One can never overemphasize its importance
- ♦ Evaluation relative to a particular language
- ♦ Are aligners something that is worthwhile to evaluate in itself?

2006-10-10

The importance of evaluation

- ♦ Evaluation is itself a first-class research activity: creation of effective evaluation methods drives more rapid progress and better communication within a research community (Hirschman, 1988:302f)
- ♦ (Johansson, 1976?)

2006-10-10

Relative to language

- ♦ Language-independence is a myth
- ♦ Good strategies for one language do not necessarily carry through to other languages
- ♦ It is important to evaluate in a particular language in context, beyond "general" evaluation

Santos (1999)

2006-10-10

Evaluation of alignment

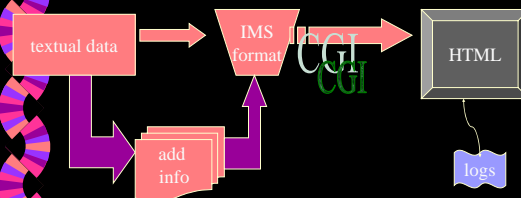
- ♦ Gaizauskas (1998:252) distinguishes between *user-visible* vs. *user-transparent* tasks : Clearly alignment is a user-transparent task
- ♦ Simard et al. (2000:38), however, states that "[bilingual text alignment] has in fact become a key element leading to recent advances in translation automation"

The COMPARA/DISPARA project

- ♦ Initiative of Ana FrankenberG-Garcia (AFG), implementation in collaboration with our project
- ♦ *COMPARA* is an English-Portuguese parallel corpus of originals and translations in the 2 directions
- ♦ *DISPARA* is a Web interface developed in connection with the *COMPARA* corpus, to make parallel corpora available through the Web

2006-10-10

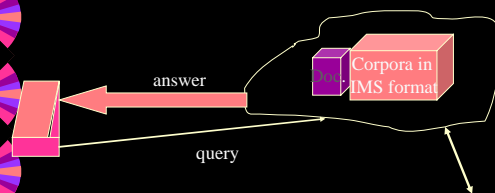
The process



Textual data ranges from simple text to SGML
Information added: from paragraph marking to parsing the text

2006-10-10

The result

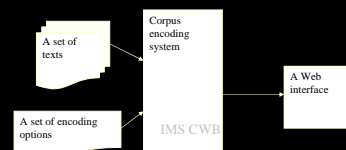


Try [http://cgi.portugues.mct.pt/acesso ...](http://cgi.portugues.mct.pt/acesso...)

Documentation on the corpora; the encoding; the kinds of queries allowed; the underlying workbench; corpus uses...

2006-10-10

Distributing corpora on the Web



2006-10-10

Creation of COMPARA

- ♦ Choice of texts and copyright clearance: AFG
- ♦ From raw to Web searchable
 - initial preparation of the texts: AFG
 - preliminary encoding and alignment: CPP
 - alignment revision and annotation: AFG
 - final encoding: CPP
- ♦ Interface (*DISPARA*) design
 - functionalities, user requirements: both
 - Web page design: AFG
 - implementation: CPP

2006-10-10

Original (?) features of DISPARA

- ♦ Inspection of translator's notes (N.T.)
- ♦ Search by alignment type (1:2, 1:1/2, etc.)
- ♦ Quantitative display of parallel distribution
- ♦ Search in the two directions
- ♦ Arbitrarily complex query string in both sides

COMPARA features that differ from ENPC

Different sizes and parts of the texts

Alignment unit always one source sentence.


Not TEI-compliant. Only <title>, <foreign>, <emph> and <note> tags

2006-10-10



DISPARA's look & feel

2006-10-10



What is alignment?

- ♦ Not as clear as one would expect...
even if narrowing to sentence alignment
- ♦ How to represent the result of alignment?
 - What counts as a sentence?
 - How to encode suppression/addition of sentences?
 - How to encode reordering of sentences?

Systems differ wildly in these respects and this has consequences for evaluation

2006-10-10




Hofland's aligner vs. EasyAlign

...from a user's perspective

♦ Human input required	♦ No human input
♦ Anchor list required	♦ No anchor list
♦ Adds <s corresp> tags	♦ Encodes in CQP format
♦ Easy to change manually the result	♦ Requires programming to change/inspect results
♦ Expects human revision	♦ Human revision implies second pass
♦ DOS	♦ Unix


Both cases are far from easy to evaluate !!!

2006-10-10



Example of TCA's output


2006-10-10



Example of EasyAlign output

Displaying alignment for [teste-i, teste] from file 'testepi.alinh'
 Enter 'h' for help.
 >> 1:1-alignment [0, 9] x [0, 111] (169)
 Six North Africans were playing boule beneath Flaubert's statue. Seis Norte-afriicanos estavam a jogar boule sob a estátua de Flaubert.
 1:1-alignment [10, 19] x [12, 22] (51)
 Clean cracks sounded over the grumble of jammed traffic. As pancadas secas ouviam-se sobre o ruído do trânsito congestionado.
 1:2-alignment [356, 390] x [386, 421] (255)
 At Trouville Flaubert's upper thigh has had to be patched, and bits of his moustache have fallen off; structural wires poke out like twigs from a concrete stub on his upper lip. Em Trouville, a parte superior da coxa de Flaubert teve de ser remendada e caíram farrapas do bigode. No lábio superior os ferros da estrutura saem como ramos de um toro de cimento.

2006-10-10



Evaluation

- ♦ First phase (Santos & Oksefjell 2000)
 - Basic quantitative measures
 - Full anchor list; full corpus
- ♦ Second phase
 - More linguistically motivated evaluation
- ♦ Comparison with other published evaluations

2006-10-10

S&O 2000's reported evaluations

1. Estimate the revision and correction involved for a given set of E-P texts
2. Estimate the contribution of the anchor word list to the above process
3. Look at the content of the anchor word list, for each entry and for each pair

2006-10-10

Interesting findings (S&O 2000)

- ♦ Many independent indicators make for a robust performance
- ♦ Cross-categorical match
 - nation.*? / ((país.*?)(naç.*?)(naciona.*?))
 - imagin.*? / imagin.*?
- ♦ Success not always linguistically motivated
 - not expected
 - wrong match; right alignment

2006-10-10

Unexpected hits (S&O 2000)

buil.* / constru.*

- *Once the job is finished the **builders** are killed*
- *Terminada a obra, os **construtores** serão mortos.*

earli.* / cedo:

- *She worked harder than anybody else, got up **earlier**, came to bed long after the others*
- *Ela trabalhava mais que qualquer um, acordava **cedo**, era a que ia dormir mais tarde.*

crim.* / crim.*:

- *... arrived in Sicily from somewhere far away, perhaps the **Crimea**, and within a few days ...*
- *... chegou à Sicília vinda de um lugar distante, talvez da **Críméia**, e em poucos dias ...*

2006-10-10

Hits by chance (S&O 2000)

look.* / olh.*

- *which makes my face **look** pallid and ill, with circles under the eyes.*
- *que dá à minha cara um aspecto pálido e doentio, com **olheiras**.*

be / ((ser))(estar)

- *one will always **be** a stranger*
- *um **ser** ('being', n.) humano sempre **será** um estranho*

couple.* / par.*

- *After a **couple** of years to settle down,...*
- *Depois de **alguns** anos **para** ('for') se estabelecer, ...*

little / pequen.*

- *was a **little** group of children, the eldest girl wheeling a pushchair with two **smaller** children,*
- *ia um grupo de crianças, a mais velha a empurrar um carrinho, com **duas** crianças mais **pequenas**, uma de cada lado*

2006-10-10

Concluding remarks (S&O 2000)

- ♦ Evaluation is a time and resource consuming task
- ♦ The strengths of an NLP system may not depend crucially on the quality of the linguistic information
- ♦ A language-dependent system may not be that language-dependent after all
- ♦ Decrease of performance by adapting to a new language pair is hard to measure
- ♦ Such a detailed study pinpoints interesting areas of textual differences between the two languages, and thus areas for contrastive research

2006-10-10

Interesting research questions

- ♦ Are grammatical or lexical clues the best anchors?
- ♦ How much does crosscategorical translation help?
- ♦ Does translation direction play a role?
- ♦ What kind of translation profile is best for alignment?

2006-10-10