

# ISA & ICA - Two Web Interfaces for Interactive Alignment of Bitexts

Jörg Tiedemann

Alfa Informatica, University of Groningen  
Oude Kijk in 't Jatstraat 26  
9712 EK Groningen, The Netherlands  
j.tiedemann@rug.nl

## Abstract

ISA and ICA are two web interfaces for interactive alignment of parallel texts. ISA provides an interface for automatic and manual sentence alignment. It includes cognate filters and uses structural markup to improve automatic alignment and provides intuitive tools for editing them. Alignment results can be saved to disk or sent via e-mail. ICA provides an interface to the clue aligner from the Uplug toolbox. It allows one to set various parameters and visualizes alignment results in a two-dimensional matrix. Word alignments can be edited and saved to disk.

## 1. Introduction

In this paper we describe two web-based interfaces for interactive alignment of parallel texts, one for sentence alignment (ISA) and one for word alignment (ICA). Both interfaces use automatic alignment as the backbone but allow users to revise the proposed alignments. The sentence aligner can also be used to create manual alignments from scratch. Manual alignments have previously been used for evaluation purposes, i.e. to produce gold standards for automatic alignment systems, see e.g. (Véronis and Langlais, 2000; Melamed, 2001). Alignment interfaces have also been shown to be effective for visualization (Smith and Jahr, 2000) and alignment improvement (Isahara and Haruno, 2000; Ahrenberg et al., 2003; Callison-Burch et al., 2004). Our interfaces include both, alignment visualization and post-editing tools for sentence and word alignment.

## 2. General Architecture

ISA and ICA are part of the Uplug toolbox for parallel text processing (Upl, 2005). They are developed to interact with other parts of the toolbox. They are implemented in PHP<sup>1</sup> and, hence, they are server-side web applications providing interfaces for alignment of parallel texts stored on the server's file system. They do not provide functions for file management (e.g. for uploading files to the server). Corpus files have to be installed and prepared off-line on the server to enable alignments via the interfaces. This can easily be done using the provided configuration routines and the Uplug tools.

Both, ISA and ICA use Uplug alignment tools for automatic alignment. Essentially they call external Uplug scripts and transform their output to information displayed in the web interfaces. ISA and ICA allow to modify alignment parameters and to edit alignment results. In this way, they can be used to visualize alignments with different settings and to post-edit alignment results.

ISA and ICA require different kinds of pre-processing. Both presume tokenized corpus data in XML format. Figure 1 shows a simple example of one tokenized sentence in

an English sample corpus.

```
<?xml version="1.0" encoding="utf-8"?>
<text>
...
<s id="s3.4">
  <w id="w3.4.1">It</w>
  <w id="w3.4.2">will</w>
  <w id="w3.4.3">be</w>
  <w id="w3.4.4">pursued</w>
  <w id="w3.4.5">with</w>
  <w id="w3.4.6">firmness</w>
  <w id="w3.4.7">and</w>
  <w id="w3.4.8">consistency</w>
  <w id="w3.4.9">.</w>
</s>
...
```

Figure 1: A simple example of tokenized input text.

Corpus files may include further markup that might be used in automatic alignment. For instance, structural markup such as paragraph boundaries can be used in the sentence aligner interface. Another example are part-of-speech tags that can be used as a feature in word alignment clues. However, once tokenized and annotated, corpus files should be static, i.e. they should not be modified anymore because ISA and ICA will produce indexes with file positions to jump to arbitrary sentences in the corpus. The indexing is done automatically when opening the web interfaces for the first time with a new corpus.

ICA (the word aligner) presumes sentence alignment and a collection of alignment clues. Sentence alignment may be done by ISA or off-line using the automatic sentence aligner integrated in Uplug. Alignment clues, however, have always to be produced off-line before starting ICA. They can also be re-used from other alignment runs. Please refer to the section on ICA for further information.

Both alignment interfaces can be used for several corpora. You may prepare any number of alignment projects using the configuration scripts. ICA and ISA will scan through the corpus directory to look for available data. Initially, the user has to select a corpus and may later always switch to another one when working with the interfaces.

<sup>1</sup>Javascript is also used for highlighting objects and tooltip windows.

ISA and ICA need to store some intermediate data to preserve settings and parameters in your session. This is done using the PHP session support that usually stores data in client-side cookies. Hence, cookies should be enabled in your browser.

Due to the server-side architecture, ISA and ICA require write permissions on the file system of the web server in the corpus directory (*corpora/*). User authentication and other routines for Internet security are not included. Access permissions and the like should be handled using other standard techniques depending on your web server software.

Furthermore, there is no multi-user mode implemented in the system. Running and editing corpus files with multiple users is no problem because intermediate data is stored in client-side cookies. However, saving data is done on the file system of the server and will cause the system to overwrite data previously saved. Storing data on disk can be disabled in the configuration scripts for each corpus to avoid this problem.

In general, ICA and ISA are not implemented for processing huge amounts of data. However, this is not the intended purpose of an interactive alignment tool.

### 3. ISA – the Interactive Sentence Aligner

The sentence aligner interface (ISA) can be used for manual and for automatic alignment. It uses a length-based approach for the automatic alignment (Gale and Church, 1993). The user can manually add or remove break points between bitext segments at any point before or after the automatic alignment. In this way (s)he can add hard boundaries to improve the automatic alignment or revise existing alignments by merging or breaking up segments.

#### 3.1. Preparing a new corpus for ISA

ISA requires two tokenized corpus files, one in a source language and one corresponding file in a target language. These files have to be stored in XML using simple markup conventions. Figure 1 shows a basic example of such a file using `<s>` tags for marking sentence boundaries and `<w>` tags for marking words. Sentences have to include unique IDs in the `id` attributes of each sentence start tag. Words should also include unique IDs to make the word aligner work correctly. Besides of these basic requirements, other markup can be included as well. Figure 2 shows the same sentence as in figure 1 but with additional markup such as phrasal chunks and part-of-speech information. It also includes another tag for *structural markup*, the paragraph boundary tag `<p>`. Structural tags can be used for improving automatic sentence alignment as we will see later on.

Finally, corpus files have to be specified in a configuration file used by ISA. For convenience, simple ISA setup including basic markup (basic XML and tokenization) and configuration can be done using a pre-defined Makefile in one go. You only need two plain text files, one containing the source language document and one containing the translation. Pre-processing and configuration will be done automatically.

```
<?xml version="1.0" encoding="utf-8"?>
<text>
<p id="1">
...
<s id="s3.4">
  <chunk id="c3.4-1" type="NP">
    <w id="w3.4.1" pos="PRP">It</w>
  </chunk>
  <chunk id="c3.4-2" type="VP">
    <w id="w3.4.2" pos="MD">will</w>
    <w id="w3.4.3" pos="VB">be</w>
    <w id="w3.4.4" pos="VBN">pursued</w>
  </chunk>
  <chunk id="c3.4-3" type="PP">
    <w id="w3.4.5" pos="IN">with</w>
  </chunk>
  <chunk id="c3.4-4" type="NP">
    <w id="w3.4.6" pos="NN">firmness</w>
    <w id="w3.4.7" pos="CC">and</w>
    <w id="w3.4.8" pos="NN">consistency</w>
  </chunk>
  <w id="w3.4.9" pos=".">.</w>
</s>
</p>
```

Figure 2: ISA input with additional markup.

#### 3.2. Handling Break Points

ISA works on the currently displayed portion of the bitext only. If the user presses the 'align' button this portion will be sent to the external sentence aligner using the break points defined in that portion as hard boundaries. ISA adds new break points to the corpus according to the automatic sentence alignment.

The user may go back and forth through the corpus and work on different portions of the text. Break points will always be stored for the entire corpus and may be saved at any time. The user may also set the amount of data to be shown (10, 20, 50 or all sentences).

All break points (automatically or manually created ones) can be revised. This is done by simply clicking on the sentence following the break point to be added or removed. These sentences will be marked with colors indicating the action that will be executed when clicking on the sentence. A light-green markup indicates addition of a new break point and a light-red markup indicates the removal of a break point (resulting in merging two segments). Sentences surrounding a newly added or removed break point will be marked after the action to make it easier to recognize the location of the recent change.

Break points can only be edited in one language at a time. It has an immediate effect on all proceeding alignments. Adding break points pushes segments in that language down to be aligned with preceding segments in the other language and removing break points pulls segments up to be aligned to previous segments. This is an efficient way to address the problem of follow-up errors which are common in automatic sentence alignment. Sometimes, however, it may cause new follow-up errors that have to be fixed manually. Usually, this can be done by adding/removing break points in the other language accordingly.

### 3.3. Using Structural Markup as Hard Boundaries

Break points can be added using structural markup from the XML sources. For instance, they can be added at each paragraph boundary in both languages. The user may add new break points incrementally using any structural markup found in the corpus. This is done by selecting the corresponding XML tag from a pull-down menu. Previous break points will be kept. For instance, a user may start by adding break points at section boundaries (if they are marked in *both* XML documents). Then, (s)he can then add break points at all paragraphs boundaries to the current set of break points. After doing so, the user may call the automatic sentence aligner which will use all the current break points as hard boundaries (section and paragraph boundaries).

The user may also reset the corpus by selecting one of the structural markup tags to be used as hard boundary and by pressing the 'reset' button. This will erase all break points added previously. There is no undo function implemented to restore previous alignments. However, intermediate results can be saved to disk and these alignments can be loaded using the 'link' tag as structural markup and clicking on the 'reset' button.

### 3.4. Using Cognates to Find Hard Boundaries

ISA includes a feature to add hard boundaries automatically using cognates as anchor points. The system looks through the bitext using a sliding window and adds hard boundaries at sentence pairs that include identical words in source and target language. The system iteratively goes through the window incrementing the distance between source and target sentence step by step. Once a new break point is identified a new sliding window is initialized. The interface provides parameters for setting the minimum length of tokens to be compared and for defining the size of the sliding window (i.e. the maximum distance between the position of the source sentence and the position of the target sentence relative to the starting point of the sliding window).

### 3.5. Alignment Results

Sentence alignment can be saved to disk at any time (if this function is not disabled in the configuration script). It will be stored in a corpus specific sentence alignment file independent from the actual corpus files. However, there is only one alignment file per corpus, hence this will be overwritten each time a new sentence alignment is saved to disk. Alignments are stored in XCES Align format using external links to sentence IDs. The advantage of this is that it avoids redundant information being stored especially in cases where one text is aligned to several target languages. An example is shown in figure 3. This format is also used as one of the input files for the word aligner.

Sentence alignment results can also be sent via e-mail in three different formats. Besides XCES Align, the system supports simple TMX and plain text formats. Figure 3 shows a small example of the different alignment formats using sentences from the Swedish Declarations of Government Policy translated into English and French.

### XCES Align

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE cesAlign
  PUBLIC "-//CES//DTD XML cesAlign//EN" "">
<cesAlign version="1.0" fromDoc="corpora/1988en.xml"
  toDoc="corpora/1988fr.xml">
<linkGrp targType="s" fromDoc="corpora/1988en.xml"
  toDoc="corpora/1988fr.xml">
<link id="SL5" xtargets="s3.4;s3.4" />
<link id="SL6" xtargets="s4.1 s4.2;s4.1" />
```

### TMX

```
<?xml version="1.0" encoding="UTF-8" ?>
<tmx version="1.4">
<header creationdate="200602161729"
  creationtool="Interactive Clue Aligner (ICA)"
  datatype="unknown" />
<body>
  <tu changedate="200602161729">
    <tuv xml:lang="en"><seg id="SRCs3.4">It will be
      pursued with firmness and consistency .</seg></tuv>
    <tuv xml:lang="fr"><seg id="TRGs3.4">Elle sera
      poursuivie avec énergie et cohérence .</seg></tuv>
  </tu>
  <tu changedate="200602161729">
    <tuv xml:lang="en"><seg id="SRCs4.1">Our policy of
      neutrality is underpinned by a strong defence .</seg>
    <seg id="SRCs4.2">That safeguards our independence .
    </seg></tuv>
    <tuv xml:lang="fr"><seg id="TRGs4.1">La politique de
      neutralité s ' appuie sur une défense forte pour la
      sauvegarde de notre indépendance .</seg></tuv>
  </tu>
```

### plain text

```
(src) [s3.4] It will be pursued with firmness and
consistency .
(trg) [s3.4] Elle sera poursuivie avec énergie et
cohérence .
-----
(src) [s4.1] Our policy of neutrality is underpinned
by a strong defence .
(src) [s4.2] That safeguards our independence .
(trg) [s4.1] La politique de neutralité s ' appuie
sur une défense forte pour la sauvegarde
de notre indépendance .
-----
```

Figure 3: An example of sentence alignment output in three different formats.

## 4. ICA – the Interactive Clue Aligner for Word Alignment

The word aligner interface (ICA) is based on the clue alignment approach (Tiedemann, 2003). Several resources (containing so-called "alignment clues") can be used to find the links between words and phrases. The user can freely select between available resources and (s)he may weight them according to their importance. Alignment clues have to be collected before running the alignment interface. This can be done using the word alignment package included in Uplug (Upl, 2005). This package produces several types of alignment clues based on co-occurrence statistics, string similarity, statistical alignment using GIZA++ (Och et al., 1999), and bootstrapped clues based on previous alignments.

### 4.1. Preparing a New Corpus for ICA

First of all, the corpus has to be sentence aligned. This can be done using ISA or the off-line tools from Uplug or any

other sentence alignment program. Sentence alignment has to be stored in XCES Align format according to the input requirements of the clue aligner (see the example in figure 3). Secondly, we need to create alignment clue databases that will be used for the automatic word alignment. The easiest way to do that is to apply the clue aligner itself which will create several clue databases from the corpus that can directly be used by ICA. Existing clue databases (for instance from previous runs) can also be used. However, be aware that they have to be appropriate for the language pair and corpus under consideration. Finally, all information about the corpus files and alignment clues have to be set in a configuration file that is used by ICA. All steps above can be taken by a simple call to a pre-defined Makefile that will perform all tasks and setups in one go using Uplug and its alignment tools.

#### 4.2. Automatic Alignment

The external word aligner is called when clicking on the 'align' button. Alignment is done for one bitext segment (a pair of aligned sentences from the sentence aligned corpus). Bitext segments can be selected from the selection box with segment IDs below the clue parameter settings. ICA sends the current sentence pair together with the selected parameters to the word alignment program and waits for the result. There is a time-out of five seconds defined to reduce the load on the server. Word alignments are shown in a two-dimensional matrix with indications of alignment certainties. These scores are computed by combining the weighted clues according to the parameter settings. Alignments are highlighted in red and other cells in the link matrix are shaded according to their clue scores. Individual clues and their scores that have been used to calculate these scores can be inspected when moving the mouse pointer over numbers in the link matrix. A tool-tip window will appear with a list of clues used. In case of aligned words (red markup) the corresponding word alignment will be included in this window as well. Word alignment pairs are also shown in a separate table beside or underneath the link matrix. These pairs will be highlighted if the mouse is moved over according cells in the link matrix.

Several parameters can be set to adjust the word aligner. Available alignment clues can be enabled or disabled. Each of them can be weighted using pull-down selection boxes attached to the clues. Furthermore, several strategies are implemented to find the optimal alignment between words and word groups according to the clue scores in the matrix (Tiedemann, 2004). A general alignment score threshold can also be set.

Finally, automatic word alignment results can be saved to disk or edited manually. For convenience there are links for aligning the next or the previous bitext segment (without saving the current alignment) and for saving the current alignment and immediately aligning the next segment.

#### 4.3. Manual Alignment

ICA as described in the previous section can be used to visualize automatic word alignment with different settings. Besides this feature, ICA also allows manual revision of automatic alignment. This is simply done by clicking on

corresponding fields in the link matrix. For convenience source and target language words and existing alignments are highlighted according to the movements of the mouse over the link matrix. A word-to-word link (marked with red background color) can be removed by clicking once on the corresponding table cell.<sup>2</sup> Adding is done a similar way. A word to word link is added by clicking once on a non-linked pair of words.<sup>3</sup> All revisions are temporary and will be lost when moving to a new bitext segment or closing the interface. However, alignment results can be saved at any time. Note that there is only one alignment file per bitext segment which will be overwritten immediately when saving new data. ICA also stores manual alignment actions (addition and removal of links). We would like to use this information later for some kind of machine learning to improve alignment while doing it. This, however, is not implemented yet.

#### 4.4. Storing Alignment Results

Word alignment (automatic or manually revised) can be stored on disk (server side). The segment ID is used as file name, i.e. there is one alignment file per bitext segment in each parallel corpus. Saving new alignments will immediately overwrite old ones from the same bitext segment. We could develop a revision tool but this is outside the scope of the current implementation. Word alignments are stored in XML using pointers to external files. Figure 4 shows an example of a word alignment result stored by ICA.

```
<link xtargets="s3.4;s3.4" id="SL5">
<wordlink xtargets="w3.4.1;w3.4.1"
lexPair="It;Elle" />
<wordlink xtargets="w3.4.2+w3.4.3;w3.4.2"
lexPair="will be;sera" />
<wordlink xtargets="w3.4.4;w3.4.3"
lexPair="pursued;poursuivie" />
<wordlink xtargets="w3.4.5;w3.4.4"
lexPair="with;avec" />
<wordlink xtargets="w3.4.6;w3.4.5"
lexPair="firmness;énergie" />
<wordlink xtargets="w3.4.7;w3.4.6"
lexPair="and;et" />
<wordlink xtargets="w3.4.8;w3.4.7"
lexPair="consistency;cohérence" />
<wordlink xtargets="w3.4.9;w3.4.8"
lexPair=".;." />
</link>
```

Figure 4: An example of a word alignment file.

#### 4.5. Inspecting Alignment Clues

ICA includes the possibility of browsing through the databases of word alignment clues. Each available clue

<sup>2</sup>Note that some links are included in larger alignment units and, therefore, removing one link does not necessarily alter the alignment immediately. Sometimes several word-to-word links have to be removed to change one unit.

<sup>3</sup>Note that adding word to word links may cause link overlaps in the source or/and the target language which will result in aligned multi-word units. ICA will add word to word links within these aligned units even if they have not been linked explicitly.

type can be opened and inspected. Figure 5 shows a screenshot when browsing through GIZA++ clues (*gw*).

source	target	score
attention	attention	1
ahead	produit	1
initiative	initiative	0.998835
give	le	0.00322263
Afghanistan	Afghanistan	1
air	réduction	1
apartheid	apartheid	1
parties	parties	1
agreement	accord	0.666667
Sweden	Suède	1
NULL	la	0.070872
between	entre	0.429406
examined	autorisations	1
breaking	mettre	0.335501
Swedish	ses	0.777001
.	.	0.702735
available	assurée	1
Proposals	propositions	0.204361
Parental	assurance	1
Later	ultérieur	1
Timrå	Timrå	1
welcomed	est	0.0296096
personnel	mieux	0.100658
personnel	personnel	0.15791
pleased	annonce	1

Figure 5: Browsing through a clue database

Each clue is a pair of source and target features (the surface strings in the example in figure 5) with an attached link score. It is possible to search for source and target patterns (using regular expressions). ICA applies for this external tools and defines again a time-out to reduce the server load. Searching and browsing through clue databases (especially big ones) is not stable mainly because of these time-outs.

## 5. Conclusions and Future Work

ICA and ISA are two web interfaces for interactive sentence and word alignment. They use the Uplug toolbox for automatic alignment and provide flexible interfaces for revising links. The sentence alignment interface ISA includes several features for improving automatic alignment such as cognate filters and possibilities of using structural markup. The word alignment interface ICA uses the clue alignment approach for automatic alignment. It provides a visualization of the alignment result and possibilities to edit links. ICA allows one to set various parameters such as clue weights and alignment strategies. Both, ISA and ICA allow one to store alignment results on disk. Sentence alignment results can also be sent via e-mail in three different formats. In the future we would like to add several features especially to the word alignment interface. We would like to include a feature to go back to previously saved alignments to

continue editing them. We would also like to learn automatically from manual corrections to improve the alignment dynamically. It could also be nice to be able to add alignment clues directly from the interface. Also, alignment at other segmentation levels (e.g. chunks/phrases) would be an interesting option.

## 6. Availability

Both interfaces are implemented in PHP with some additional Javascript calls. They are freely available from the Uplug project at <http://sourceforge.net/projects/uplug/>

## 7. References

- Lars Ahrenberg, Magnus Merkel, and Michael Petterstedt. 2003. Interactive word alignment for language engineering. In *Conference Companion of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 49–52, Budapest, Hungary.
- Chris Callison-Burch, Colin Bannard, and Josh Schroeder. 2004. Improved statistical translation through editing. In *European Association for Machine Translation (EAMT) Workshop*, University of Malta, Valletta, Malta.
- William A. Gale and Kenneth W. Church. 1993. A program for aligning sentences in bilingual Corpora. *Computational Linguistics*, 19(1):75–102.
- H. Isahara and M. Haruno. 2000. Japanese-english aligned bilingual corpora. In Jean Véronis, editor, *Parallel Text Processing: Alignment and Use of Translation Corpora*. Kluwer Academic Publishers, Dordrecht.
- Dan I. Melamed. 2001. *Empirical Methods for Exploiting Parallel texts*. The MIT Press, Cambridge, MA.
- Franz Josef Och, Christoph Tillmann, and Hermann Ney. 1999. Improved alignment models for statistical machine translation. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC)*, pages 20–28, University of Maryland, MD, USA.
- Noah A. Smith and Michael E. Jahr. 2000. Cairo: An alignment visualization tool. In *Proceedings of The Second International Conference on Language Resources and Evaluation (LREC)*, pages 549–552, Athens, Greece.
- Jörg Tiedemann. 2003. *Recycling Translations – Extraction of Lexical Data from Parallel Corpora and their Application in Natural Language Processing*. Ph.D. thesis, Department of Linguistics, Uppsala University, Uppsala/Sweden.
- Jörg Tiedemann. 2004. Word to word alignment strategies. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*, pages 212–218, Geneva, Switzerland.
2005. Uplug: Tools for linguistic corpus processing, word alignment and term extraction from parallel corpora. available at <http://sourceforge.net/projects/uplug>.
- Jean Véronis and Philippe Langlais. 2000. Evaluation of parallel text alignment systems. the ARCADE project. In Jean Véronis, editor, *Parallel Text Processing*, Text, speech and language technology series, chapter 19. Kluwer Academic Publishers, Dordrecht, August.

## Interactive Sentence Alignment

[10 | 20 | 50 | all | next page >>](#)

[Help?](#)

s1.1	Statement of Government Policy by the Prime Minister , Mr Ingvar Carlsson , at the Opening of the Swedish Parliament on Tuesday , 4 October , 1988	Declaration de Politique Générale du Gouvernement présentée mardi 4 octobre 1988 devant le Riksdag par Monsieur Ingvar Carlsson , Premier Ministre .	s1.1
s2.1	Your Majesties , Your Royal Highnesses , Mr Speaker , Members of the Swedish Parliament	Majestés , Altesse s Royales , Monsieur le Président , Mesdames et Messieurs les députés !	s2.1
s3.1	Sweden' s policy of neutrality is of decisive importance for our peace and independence	La politique suédoise de neutralité revêt une importance capitale pour la paix et l ' indépendance de notre pays .	s3.1
s3.2	It also contributes to stability and détente in our part of the world	Elle contribue également à la stabilité et à la détente dans notre secteur du monde .	s3.2
s3.3	There is wide popular support for this policy	Cette politique recueille une large adhésion populaire .	s3.3
s3.4	It will be pursued with firmness and consistency	Elle sera poursuivie avec énergie et cohérence .	s3.4
s4.1	Our policy of neutrality is underpinned by a strong defence	La politique de neutralité s ' appuie sur une défense forte pour la sauvegarde de notre indépendance .	s4.1
s4.2	That safeguards our independence	Il ne sera jamais toléré de violations du territoire suédois .	s4.2
s4.3	Violations of Swedish territory will never be accepted	L ' armée va être réformée et rendue plus efficace .	s4.3
s4.4	The army will be reorganized with the aim of making it more effective	Le Gouvernement a la ferme résolution de rechercher des solutions de large consensus sur les questions importantes pour notre sécurité nationale .	s4.4
s4.5	It is the Government' s intention to seek broad solutions in issues that are of importance for our national security		
s5.1	The Government welcomed the agreement between the United States and the Soviet Union on the elimination of land- based intermediate- range nuclear weapons	Le Gouvernement s ' est félicité de l ' accord conclu entre les Etats- Unis et l ' Union Soviétique sur l ' élimination des missiles nucléaires à portée intermédiaire basés au sol .	s5.1
s5.2	Efforts must now , among other things , aim at reducing strategic arms and bringing about a comprehensive test ban treaty	Les efforts doivent maintenant viser notamment à réduire les armements stratégiques et à mettre au point un accord sur l ' arrêt complet des essais nucléaires .	s5.2
s5.3	Sweden is acting in support of this inter alia within the framework of the Six Nation Initiative	La Suède y travaille , en particulier dans le cadre de l ' Initiative des Six Nations .	s5.3

Figure 6: The Interactive Sentence Aligner ISA

## Interactive Clue Alignment

[Help?](#)

clue weight clue weight clue weight clue weight clue weight

dice 0.01  gw 0.01  dl3x 0.01  dlx 0.01  enfrpp 0.01

sim 0.01  gwi 0.01  dl 0.01  enfrp 0.01

SL1      [\[ save & align next \]](#) [\[ align next \]](#) [\[ align previous \]](#)

[1] Your Majesties , Your Royal Highnesses , Mr Speaker , **Members** of the Swedish Parliament .

[2] Majestés , Altesse s Royales , Monsieur le Président , Mesdames et **Messieurs** les députés !

	Majestés	,	Altesse s Royales	,	Monsieur le Président	,	Mesdames et Messieurs	les députés	!		source	target	
Your	14.95	0	5	0	0	0	0	0	0	0	Your Majesties	Majestés	
Majesties	17.7	0	15.5	0	0	4.44	0	0	0	5.56	0	6.67	
,	0	22.57	0	0	22.57	0	4.07	0	22.57	0	4.77	0	3.28
Your	14.95	0	5	0	0	0	0	0	0	0	0	0	
Royal	0	0	10	7.14	0	2.5	0	0	0	0	0	0	
Highnesses	0	0	5	19.9	0	0	0	0	0	4	0.05	5	
,	0	22.57	0	0	22.57	0	4.07	0	22.57	0	4.77	0	3.28
Mr	0	0	0	0	0	24.8	0	0	0	0	0	0	
Speaker	0	0	0	0	0	0.1	19.9	0	0	0	0	0	
,	0	22.57	0	0	22.57	0	4.07	0	22.57	0	4.77	0	3.28
Members	0	0	0	0	0	0	0	0	0	26.03	0	15.5	
of	0	4.36	0	0	4.36	0	4.06	0	4.36	0	4.75	0.0	
the	0	4.07	0	0	4.07	0	11.72	0	4.07	0	4.47	0	
Swedish	0	0	0	0	0	4.44	0	0	0	0	0	0	
Parliament	0	0	0	0	0	4.44	0	6	0	0	0	0	
,	0	4.22	0	0	4.22	0	4.2	0	4.22	0	6.26	0	4.95

type	source	target	score
giza-word	Members	Messieurs	0.01
string similarities	members	messieurs	0.005555555555555555

All values X 1000!

Figure 7: The Interactive Clue Aligner ICA