

TC-STAR 2006 End User Agreement
Access to the data for the OpenLab_2006 Training initiative

INSTRUCTIONS for COMPLETION

Please compile and submit the end-user agreement in two original signed copies (this page is NOT part of the agreement).

Please complete the reference:

(Reference: ELDA/TC-STAR2006/Openlab_2006/name_of_your_institution)

at the top of page 1 of the contract with the name of your institution.

Each page of the agreement should also be initialed (the person signing the agreement should put his/her initials on each and all pages).

The End-User named must be a legal institution, or a department or section of a named legal institution, not an individual nor a project - the person signing this Agreement should be the person authorized by the institution for such signatures.

In Exhibit A, please tick the boxes for the language resources you are requesting.

In Exhibit B, you must insert the address of the location where the data will be used.

In Exhibit C, indicate clearly which tasks you will be performing. You are only entitled to have access to data for these tasks. If, due to circumstances beyond your control, you are unable to complete any or all of these tasks, you should inform us.

If you later decide that you would like to perform an additional task, please inform us and we will send you an addendum to the agreement if additional data is needed.

Exhibit E is a legal statement that should be send back once the Openlab_2006 is over and the data is completely deleted from all your computers, archiving media, etc.

All forms must be sent by mail to:

Christelle Ayache
ELDA
55-57 rue Brillat-Savarin
75013 Paris

France

To speed up the delivery process, this agreement can be faxed to

Christelle Ayache +33 1 43 13 33 30

and sent by regular mail afterwards.



Technology and Corpora for Speech to Speech Translation

LANGUAGE RESOURCES

END-USER AGREEMENT

(Reference: ELDA/TC-STAR2006/OpenLab_2006/ *Please Insert Name of End-User here*)

This agreement is made by and between:

.....
(hereinafter called END-USER), having its principal place of business at:

.....
AND

ELDA S.A., the distribution agency commissioned by ELRA, (acting on behalf of the TC-STAR Project¹, and hereinafter called TC-STAR), having its principal place of business at:55-57 rue Brillat Savarin - 75013 Paris, FRANCE

registered at the Tribunal de Commerce de Paris : RCS Paris B 402 781 876 (95b147 95)

Preamble:

TC-STAR-OpenLab 2006 (herein referred to as OpenLab_2006) is a training initiative of the European Integrated Project TC-STAR, Technologies and Corpora for Speech-to-speech Translation Research. OpenLab 2006 aims to expand the TC-STAR research community in the areas of Automatic Speech Recognition (ASR) and Spoken Language Translation (SLT).

Students and young researchers in these areas are invited to contribute on shared TC-STAR project tasks.

The translation of European Parliament speeches from Spanish to English is the application domain of interest. Contributions on the following and other closely related topics will be welcome:

- Integration of ASR and SLT
- Statistical Models for SLT
- System combination in ASR and SLT
- Morphology and Syntax in SLT
- Error analysis in SLT

Several months before the meeting in Trento, language resources (object of this agreement) and tools will be made available to interested participants for the exclusive use within the TC-STAR Openlab2006 and during the

¹ The term 'TC-STAR Project' refers to the TC-STAR Integrated Project documented at <http://www.tc-star.org/>

period up to the completion of the Openlab_2006.. Word graphs and n-best lists generated by different ASR and SLT systems will be provided, as well as training and testing collections to develop and evaluate a SLT system.

Participants will present and discuss their results in Trento, and will have the opportunity to attend tutorial speeches held by experts. Participation in OpenLab 2006 is free.

whereby it is agreed as follows:

1. **Language Resources**, object of this Agreement, for which ELDA or the TC-STAR partners obtained distribution rights from the rightful holders (hereinafter called PROVIDERS) for the use within the TC-STAR OpenLab2006 campaigns, are described in Exhibit A.
2. The site of use of the **Language Resources** is mentioned in Exhibit B.
3. Within this Agreement ELDA grants END-USER, engaged in *bona fide* language engineering research, the non-exclusive right to use the **Language Resources**, exclusively for the purposes defined in the OpenLab_2006 training. **ELDA grants the END-USER the rights to use the data for a limited period of time and up to the end of the OpenLab 2006 training and no later than end of April 2006..**
4. END-USER is not permitted to reproduce the **Language Resources** for commercial or distribution purposes and to commercialize (or distribute for free) in any form or by any means the **Language Resources** or any derivative product or services based on all or a substantial part of it.
5. ELDA grants END-USER the right to reproduce the **Language Resources** temporarily or permanently, to translate, adapt, arrange and modify by any means the **Language Resources**, if one or all of these acts are necessary to access and make the agreed use of the contents of the **Language Resources**. END-USER is permitted to make a copy of the **Language Resources** for purposes of archiving only during the OpenLab_2006 initiative.
6. END-USER agrees to return and delete all copies (including the one made for archiving purposes) at the end of the OpenLab_2006 training initiative from all its computers, archiving media, etc. END-USER commits to send by registered mail the notification attached to this agreement duly signed by its legal representative.
7. Without prejudice to the other provisions, the rights referred to herein shall be non transferable to any other entity. The **Language Resources** shall not be transferred to or accessed from any other site.
8. END-USER acquires no ownership, rights or title in all or any parts of the **Language Resources**.
9. ELDA and TC-STAR partners and PROVIDERS accept no responsibility for the accuracy or completeness of the data or for the consequences of their use. TC-STAR partners PROVIDERS give no warranty for merchantability and/or fitness for a particular purpose of the **Language Resources**.
10. END-USER shall give appropriate references to TC-STAR AS WELL AS PROVIDERS in scholarly literature when the **Language Resources** are mentioned.
11. END-USER shall not use the name of TC-STAR nor the names of PROVIDERS in any publication in any manner that would imply an endorsement of END-USER or any product or service offered by END-USER.
12. END-USER has no right or authority to incur, assume or create, in writing or otherwise, any warranty, liability or other obligation of any kind, express or implied, in the name of or on behalf of TC-STAR, it being intended that each party shall remain an independent contractor responsible for its own actions.
13. Neither party shall be responsible for damages caused by the other party's non-fulfillment of provisions of this AGREEMENT.
14. Both parties exclude all liability of whatsoever nature for direct, consequential or indirect loss or damage suffered by the other.
15. TC-STAR agrees to provide **Language Resources** to END-USER FREE OF CHARGE.

16. Signature of the present agreement by END-USER indicates the END-USER's intention to participate in the OpenLab_2006 initiative. If, for whatever reason, END-USER can not participate, this agreement shall be automatically terminated and article 6 clauses shall immediately apply.
17. END-USER will maintain and post a list of people with current and recently-terminated access to the **Language Resources**. END-USER shall maintain such appropriate records and ELDA shall have the right to audit such records upon reasonable prior notice using an independent auditor of its choice (the "Auditor"). The Auditor shall be bound to keep confidential the details of the business affairs of END-USER.
18. END-USER agrees not to disclose nor communicate to anyone any information concerning or contained in technical or commercial documents relating to the Language Resources placed at END-USER's disposal by ELDA, except to its officers, directors or employees on a "need-to-know" basis.
19. In order to preserve the desired, pre- competitive nature of the TC-STAR conferences, TC-STAR consortium has developed a set of guidelines constraining the dissemination and publication of TC-STAR OpenLab2006 results. These guidelines are meant to preclude the publication of incomplete or inaccurate information that could damage the reputation of the conference or its participants and could discourage participation in future conferences. The guidelines (listed in Exhibit D) shall be implemented by END-USER.
20. This Agreement is subject to, construed and interpreted in accordance with the Law of France. Should it not be possible to settle amicably differences of interpretation out of this Agreement, then the case shall be brought before the regular courts of law for a decision. The "Tribunal de Commerce de Paris" shall be the only competent court. This AGREEMENT may be terminated by ELDA if END-USER breaches any material term or provision of this AGREEMENT, including article 18. In the event that one of these articles is considered by French law or a French court decision as null, this invalid or illegal article shall not affect validity, legality and enforceability of the remaining provisions.

The entire AGREEMENT is composed of the 20 articles herein together with Exhibits A, B, C and D thereafter.

In witness whereof, intending to be bound, the parties hereto have executed this AGREEMENT by their duly authorized officers.

AUTHORISED BINDING SIGNATURES:

 On behalf of
 Name:
 Title:
 Date:

 On behalf of the TC-STAR Consortium
 Name: Khalid CHOUKRI
 Title: ELDA Managing Director
 Date:

EXHIBITS

Exhibit A: Language Resources

Language Resources refer to the resources selected below (*please tick the resources you are requesting*):

- EPPS English/Spanish SLT training data set:** Parallelised EPPS Final Text Editions for 1996-2004 – Copyright European Communities Official Publication Office and RWTH.
- EPPS Spanish-to-English SLT development and test sets:** Transcriptions of EPPS from 26 October to 18 November 2004 in English and Spanish – Copyright ELDA – Final Text Edition (Spanish) of EPPS from 26 October to 18 November – Copyright European Communities Official Publication Office – Reference translations of Spanish EPPS Transcriptions into English – Copyright ELDA – Reference translations of Spanish Final Text Edition into English – Copyright ELDA.
- SLT scoring tools** Implementation of the following metrics BLEU/NIST v11a, mWER: multiple reference word error rate, mPER: multiple reference position-independent word error rate, mCER: multiple reference character error rate, WNM: Weighted N-gram Model – Copyright ELDA, IBM, University of Leeds.

Exhibit B: SITE OF USE: (insert here below the address of the location where the data will be used)

.....

.....

.....

.....

EXHIBIT C: Guidelines constraining the dissemination and publication of TC-STAR OpenLab2006 results are:

- a) **SCIENTIFIC OR TECHNICAL PUBLICATIONS:** Scientific or technical publications, including newsletters from universities or research laboratories, should adhere to community standards for fairness and objectivity and should accurately and clearly state the limitations of the testing conditions and other factors which might influence scores. The experimental nature of the tasks, data and evaluation procedures should also be stated. The full conference proceedings should always be referenced.
- b) **ADVERTISEMENTS:** No advertisements using the TC-STAR OpenLab2006 results can be placed in magazines, journals, newspapers, or other publications.
- c) **PRESS RELEASES:** Press releases about TC-STAR results to organizations with national/international coverage are also prohibited.
- d) **MARKETING LITERATURE, LOCAL NEWSLETTERS:** Although it is recognized that extensive evaluation discussions are not appropriate in this type of literature, it is expected that any claims made on the basis of evaluation results are accurate, that the evaluation measures used to substantiate these claims are stated, and that a reference is made to the full conference proceedings. Where promotional material is subject to prepublication revision by the media, the author should make every effort to see that the revision does not cause a violation of the guidelines.
- e) **CROSS-SYSTEM COMPARISONS:** Cross-system comparisons may not be made with other named teams for individual tests, and may only be made when they are supported by accepted methods of statistical significance testing. Comparisons must be accompanied by the results of those tests and should reference the publication of those tests in the conference proceedings or related literature. Informal, qualitative comparisons with recognized baselines or benchmarks, and with general levels or trends in performance, must be clearly stated to be such and thus open to statistical reassessment.
- f) **PUBLICATION WITH TC-STAR REFERENCES:** A copy of any publication that quotes or contains references to the TC-STAR OpenLab2006 results must be provided to the TC-STAR coordinator.

EXHIBIT D: Notification letter to return to ELDA after the termination of this contract

To:

Khalid CHOUKRI
ELRA/ELDA
55 Rue Brillat-Savarin,
F-75013 Paris, France

I hereby, acting as the legal representative of (*Insert the name of your institution*) certifies that my institution has deleted all copies of the data received within the OpenLab_2006 training initiative. I declare that we have neither copies nor archived copies of the above mentioned data. I declare that we have not kept any analyses or interpretations of the linguistic properties of the information that allow reconstructing or exploiting the Language resources object of this agreement (e.g. signal analysis).

AUTHORISED BINDING SIGNATURE:

On behalf of:
Name:
Title:
Date²:

² (no later than 4 weeks after the end of OpenLab_2006)