

# **TC-STAR 2007 End User Agreement**

## **INSTRUCTIONS for COMPILATION**

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-STAR2007/xxx) 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, 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.

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

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

All forms must be sent by mail to:

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



# Technology and Corpora for Speech to Speech Translation

## LANGUAGE RESOURCES

### END-USER AGREEMENT

(Reference: ELDA/TC-STAR2007/.....)

**This agreement is made by and between:**

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

AND

**ELDA S.A.**, the evaluation and language resources distribution agency commissioned by ELRA, (acting on behalf of the TC-STAR Project<sup>1</sup>, 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)

**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 evaluation campaigns, are described in Exhibit B.
2. The site of use of the **Language Resources** is mentioned in Exhibit C.
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 of system evaluation within the TC-STAR campaign for 2007.
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. Summaries, analyses and interpretations of the linguistic properties of the information may be derived and published, provided it is not possible to reconstruct the information from these summaries. Small excerpts of the information may be displayed to others or published in a scientific or technical context, solely for the purpose of describing the research and development and related issues. Any such use shall not infringe on the rights of any third party including, but limited to, the authors and publishers of the excerpts.

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

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.
6. 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.
7. END-USER acquires no ownership, rights or title in all or any parts of the **Language Resources**.
8. 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**.
9. END-USER shall give appropriate references in scholarly literature when the Language Resources are mentioned. The following acknowledgement is required: "TC-STAR 2007 Evaluation Package, distributed by ELRA on behalf of TC-STAR Consortium". Reference shall also include information on the Language Resources as defined in Exhibit B.
10. 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.
11. 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.
12. Neither party shall be responsible for damages caused by the other party's non-fulfillment of provisions of this AGREEMENT.
13. Both parties exclude all liability of whatsoever nature for direct, consequential or indirect loss or damage suffered by the other.
14. TC-STAR agrees to provide **Language Resources** to END-USER free of charge under this license.
15. Signature of the present agreement by END-USER indicates the END-USER's intention to participate in the TC-STAR system evaluation activity for the tasks stated in Exhibit A.
16. 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.
17. 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.
18. In order to preserve the desired, pre- competitive nature of the TC-STAR evaluations, TC-STAR consortium has developed a set of guidelines constraining the dissemination and publication of TC-STAR evaluation results. These guidelines are meant to preclude the publication of incomplete or inaccurate information that could damage the reputation of the project or its participants and could discourage participation in future evaluation campaigns. The guidelines (listed in Exhibit D) shall be implemented by END-USER.
19. 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 19 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: Tasks to be carried out by USER***

Indicate clearly below in which tracks/tasks you will be participating and which data collections you will be using. A table showing the collections which can be used in each track is given in the instructions on How to participate on the TC-STAR Evaluation pages website at <http://www.elda.org/en/proj/tcstar-wp4/index.htm>.

**ASR Track:** Please indicate which task(s) you intend to participate to.

- English EPPS
- Spanish EPPS + CORTES
- Chinese Mandarin VOA

**SLT Track:** Please indicate which task(s) you intend to participate to.

- English-to-Spanish EPPS
- Spanish-to-English EPPS + CORTES
- Chinese Mandarin-to-English VOA

**TTS Track:** Please indicate which task(s) you intend to participate to.

- Text Processing – English
- Acoustic Synthesis – English
- TTS System – English
- TTS System – Spanish
- TTS System – Chinese Mandarin
- Voice Conversion – English
- Voice Conversion – Spanish
- Voice Conversion – Chinese Mandarin
- Voice Conversion based on found data

***Exhibit B: Language Resources***

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

- ASR English Training data set:**
  - European Parliament Plenary Sessions (EPPS) Video Broadcast recordings in English from 3 May 2004 to 31 May 2006 – Copyright European Communities Official Publication Office – Transcriptions – Copyright RWTH-Aachen
- ASR English Development and Evaluation data sets:**
  - European Parliament Plenary Sessions (EPPS) Video Broadcast recordings in English from 6 June 2005 to 28 September 2006 – Copyright European Communities Official Publication Office – Transcriptions – Copyright ELDA
- ASR Spanish Training data set:**
  - European Parliament Plenary Sessions (EPPS) Video Broadcast recordings in English from 3 May 2004 to 31 May 2006 – Copyright European Communities Official Publication Office – Transcriptions – Copyright Universidad Polit cnica de Catalunya
  - CORTES Spanish Parliament (CORTES) Video Broadcast recordings in Spanish from July 2004 to November 2004 - Copyright Congreso de los Diputados – Transcriptions – Copyright Universidad Polit cnica de Catalunya
- ASR Spanish Development and Evaluation data sets:**
  - European Parliament Plenary Sessions (EPPS) Video Broadcast recordings in English from 6 June 2005 to 28 September 2006 – Copyright European Communities Official Publication Office – Transcriptions – Copyright ELDA
  - CORTES Spanish Parliament (CORTES) Video Broadcast recordings in Spanish from 1 December 2004 to 20 June 2006 - Copyright Congreso de los Diputados – Transcriptions – Copyright ELDA
- ASR VOA Chinese Development and Evaluation data sets:**
  - Audio recordings of Voice of America in Mandarin Chinese from 1 to 11 December 1998 and from 26 to 27 December 1998 – Copyright Voice of America – Transcriptions – Copyright ELDA
- SLT English/Spanish EPPS training data set:**
  - Parallelized EPPS Final Text Editions for 1996-2006 – Copyright European Communities Official Publication Office and RWTH.
- SLT English-to-Spanish Development and Evaluation data sets:**
  - Transcriptions of EPPS from 6 June 2005 to 28 September 2006 in English – Copyright ELDA – Final Text Edition (English) of EPPS from 6 June 2005 to 28 September 2006 – Copyright European Communities Official Publication Office – Reference translations of English EPPS Transcriptions into Spanish – Copyright ELDA – Reference translations of English Final Text Edition into Spanish – Copyright ELDA.
- SLT Spanish-to-English Development and Evaluation data sets:**
  - Transcriptions of EPPS from 6 June 2005 to 28 September 2006 in Spanish – Copyright ELDA – Final Text Edition (Spanish) of EPPS from 6 June to 28 September 2006 – 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.
  - Transcriptions of CORTES from 1 December 2004 to 20 June 2006 in Spanish – Copyright ELDA – Final Text Edition (Spanish) of CORTES from 1 December 2004 to 20 June 2006 – Copyright Congreso de los Diputados – Reference translations of Spanish Transcriptions into

English – Copyright ELDA – Reference translations of Spanish Final Text Edition into Spanish – Copyright ELDA.

- SLT Chinese-to-English Development and Evaluation set:**
  - Reference translations of the VOA Chinese transcriptions from December 1998 into English – Copyright ELDA.
- TTS Text Processing– English: Evaluation data set**
  - Text excerpts from EPPS FTE (English) from June 2006 to September 2006 – Copyright European Communities Official Publication Office.
- TTS Acoustic Synthesis – English: Evaluation data set**
  - Prosodic description (XML) made from recordings of EPPS FTE (English) paragraphs – Copyright ELDA.
- TTS System – English: Evaluation data set**
  - Set of paragraphs extracted from EPPS FTE (English) – Copyright European Communities Official Publication Office.
- TTS System – Spanish: Evaluation data set**
  - Set of paragraphs extracted from EPPS FTE (Spanish) – Copyright European Communities Official Publication Office.
- TTS System – Chinese Mandarin: Evaluation data set**
  - Set of paragraphs extracted from the TTS evaluation corpus for National High-Tech program 863 TTS evaluation in 2003. ref 2003-863-002. Copyright ChineseLDC, CIPSC (<http://www.chineseldc.org>).
- TTS Voice Conversion – English: Development and Evaluation data sets**
  - Data and Annotation for Spanish/English voice conversion – Copyright Universidad Politécnic de Catalunya
- TTS Voice Conversion – Spanish: Development and Evaluation data sets**
  - Data and Annotation for Spanish/English voice conversion – Copyright Universidad Politécnic de Catalunya
- TTS Voice Conversion – Chinese Mandarin: Development and Evaluation data sets**
  - Data and Annotation for Mandarin voice conversion (ref. TCSTTS6zh-CN) – Copyright Nokia and Siemens.
- TTS Voice Conversion based on found data: Development set**
  - Target: audio excerpts from the EPPS recordings (2 English politician speakers) – Copyright European Communities Official Publication Office.
  - Source: TTS samples provided by IBM (Spanish).
- TTS Voice Conversion based on found data: Development and Evaluation data sets**
  - TTS samples provided by IBM (Spanish).

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

.....

.....

.....

.....

***EXHIBIT D: Guidelines constraining the dissemination and publication of TC-STAR evaluation 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 evaluation 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 evaluation results must be provided to the TC-STAR coordinator.