

OSCOTT

Minimizing costs and enhancing minority language use: Open Source Collaborative Translation Technology OSCOTT

Peter Sandrini

Overview

- Introduction and motivation
- Multilingualism
- OSCOTT Framework
 - Terminology
 - Translation Management
 - Web Content Management
 - Cooperation and Exchange

Requirements for official translation

- **investment in**
 - **people**
translators training, ongoing education
 - **organizations**
setting up translation offices and departments
 - **technology**
CAT tools and translation infrastructure

FOSS for freelance translators

- USBTrans
- *tuxtrans* (tuxtrans.org)
- Free Software
- Target: individual translators
- No group functions
- No support for online cooperation
- No server features



Social translation initiatives

- Translators without borders
www.translatorswithoutborders.com
- The Rosetta Foundation
www.therosettafoundation.org
- Mondo Lingua Initiative
www.mondo-lingua.org
- Translators and Interpreters for Solidarity, ECOS
<http://ecosteis.wordpress.com>
- Babels
www.babels.org
- ...



Linguistic human rights

- European Charter for Regional and Minority Languages (EU 1992)
- Universal Declaration of Linguistic Rights (UNESCO 1996)
 - Article 8: "1. All language communities have the right to organize and manage their own resources so as to ensure the use of their language in all functions within society.
2. All language communities are entitled to have at their disposal whatever means are necessary to ensure the transmission and continuity of their language."
 - Article 11: "All language communities are entitled to have at their disposal whatever means of translation into and from other languages are needed to guarantee the exercise of the rights contained in this Declaration."
- Declaration of the Rights of Indigenous Peoples (UN 2007)



Multilingualism: India



- 415 languages listed by SIL Ethnologue
- 30 languages are spoken by more than a million native speakers
- 122 by more than 10,000
- Regional linguistic conflicts over the role of Hindi as official language

Multilingualism: South Sudan

- 133 languages in 3 language families: Afro-Asiatic, Nilo-Saharan, and Niger-Congo
- Official language: English
- Most languages in the South



Nation-building agenda

- Independence from the northern part
- Nation-building efforts:
 - education, health, infrastructure,
 - law enforcement, agriculture, economy
 -
 - language policy and infrastructure support of language diversity as an element of democratization

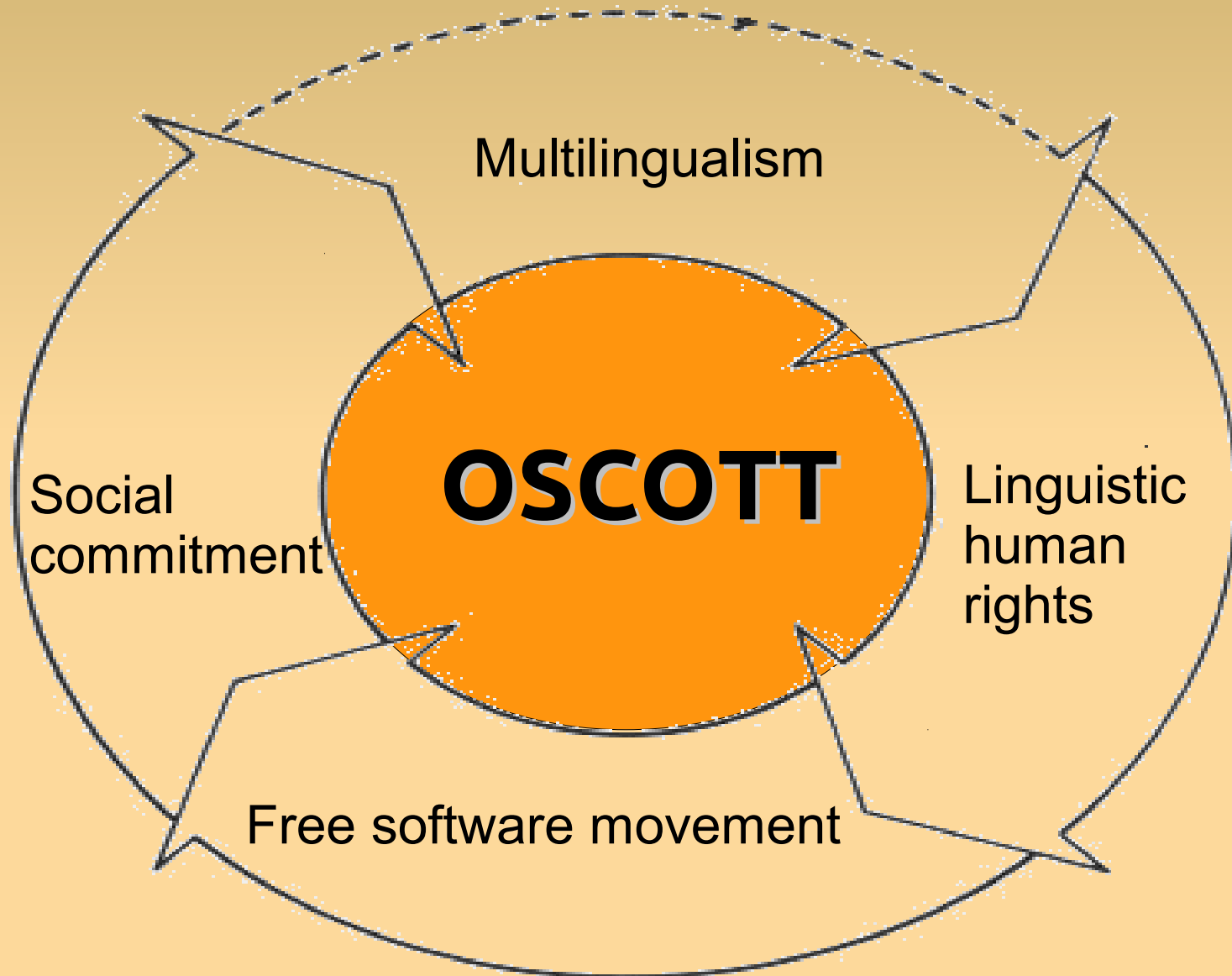


Target groups

- All organizations or institutions in charge of official multilingualism
- Local, regional and national level
- Disadvantaged communities



Motivation



Requirements

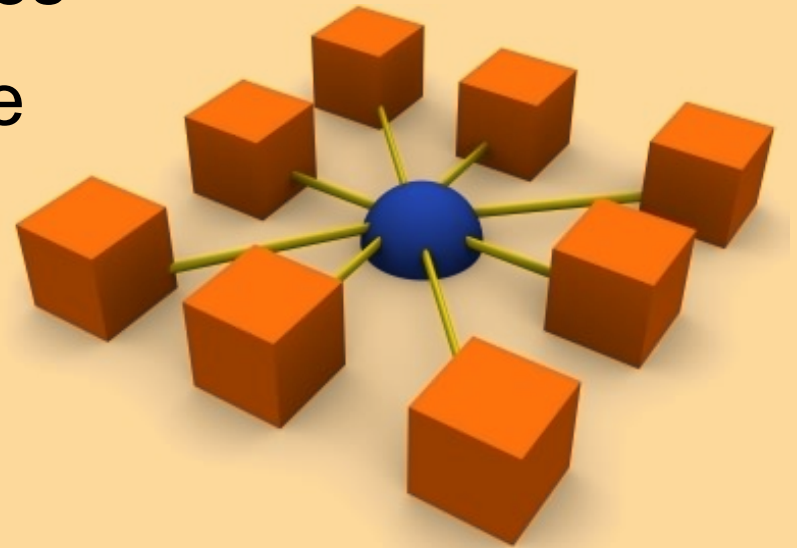
- Language policy:
 - ...
 - Terminology policy
 - Translation policy
- Language infrastructure
 - Terminology infrastructure
 - Translation infrastructure

OSCOTT aims at

- providing a free Translation Technology Infrastructure for the support of language minorities
- bundling existing free collaborative translation technology applications into one package
- accompanying this with a set of guides and standard procedures for
 - translation management
 - terminology harmonization
 - language ressources management

Translation technology infrastructure

- a TTI is a set of software applications designed to increase the efficiency of multilingualism, while improving quality and reducing costs by making previous translations and terminology available for future access and re-use
- Central language data repositories
- Internet-based with web interface



Open

- **Open Source**
- **Open Data**
- **Open Content**
- **Open Access**
- **Open Services**

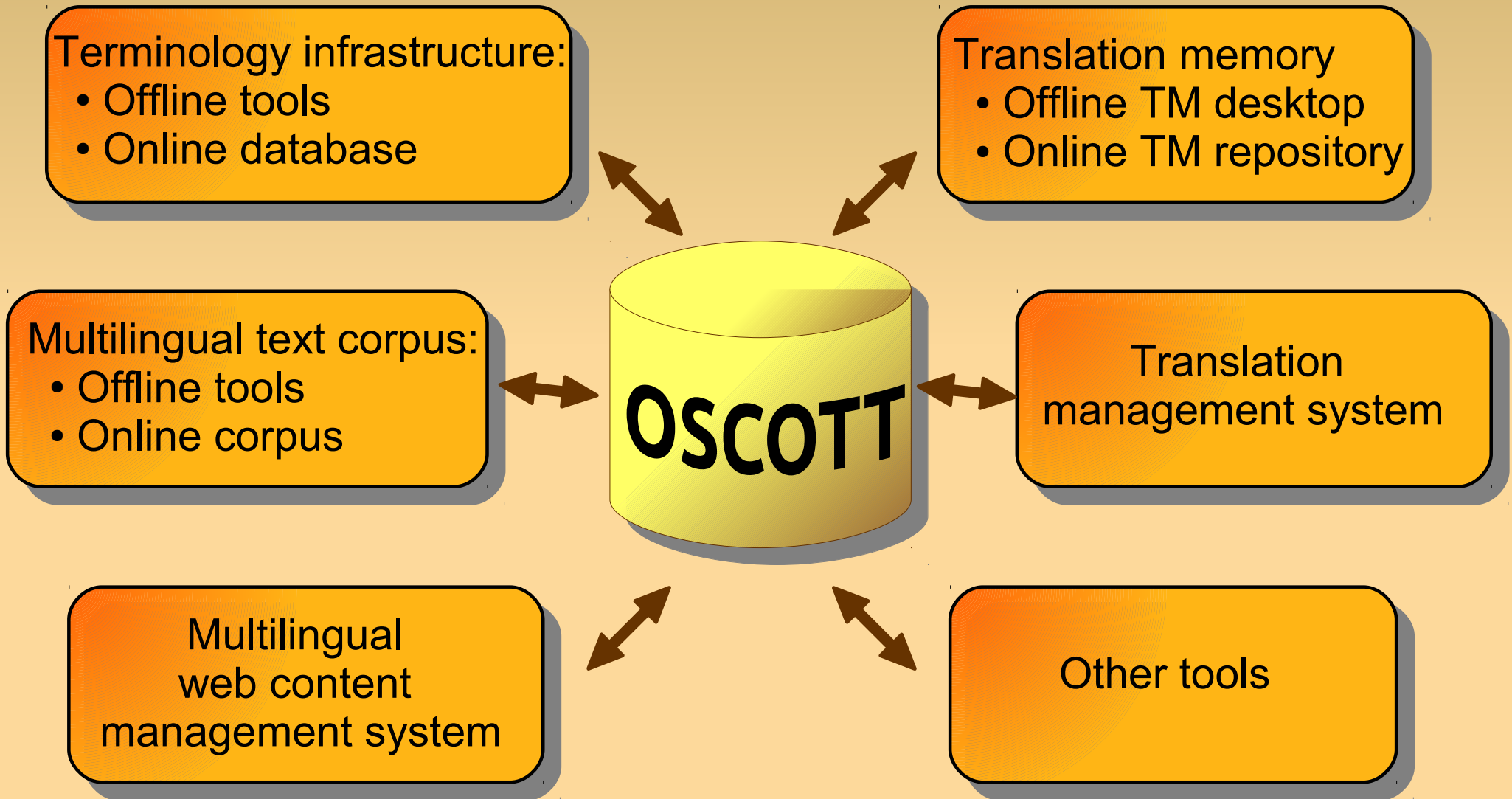
“A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike.”

<http://opendefinition.org/>

“A service is open if its source code is Free/Open Source Software and non-personal data is open as in the Open Definition.”

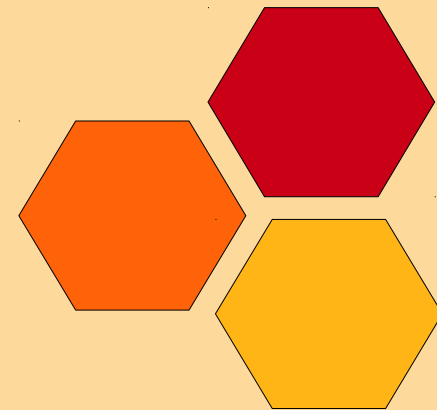
<http://opendefinition.org/>

Components



Terminology infrastructure and projects

- Project management tools
- Terminology management and database
- Terminology: main areas
 - Official institutional names
 - Education
 - Toponymy
 - Legal terminology
 - Political terminology
 - ...



Terminology Policy

Guidelines for Terminology Policies,
Formulating and Implementing
Terminology Policy in Language
Communities (UNESCO 2005):



„In this context, terminological data, their availability, and free access to these data represent one of the factors of socio-economic development”

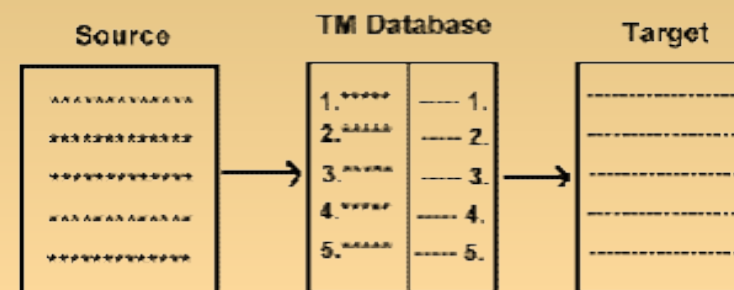
Terminology

- Offline tools
 - FD-Termbase, CmapTools, Phrase Extractor, editors,
- Online tools
 - TermWiki
 - ? (OS online collaborative TMS)



Translation memory

- Offline tool: OmegaT, alignment
- Online tool
 - TinyTM
 - Pootle
 - TuMatXa - translation memory repository
 - ...
- Online translation management:
 - GlobalSight
 -]project open[



Multilingual text corpus

- Offline tools:
 - TextSTAT
 - AntConc
- Online repository
 - ?



Multilingual web content management

- Drupal + XLIFF-Tools
- Joomla + Joomfish
- Wordpress + Qtranslate

Other tools



Research

- survey on tools in use by minorities
- list of key modules
- integration of available applications
 - interaction between tools
 - interdependencies
 - formats
- development of missing tools (e.g. OS TMS)
- documentation and guidelines



OSCOTT advantages

- Know-How on translation technology
- Know-How on management of language resources
- digital knowledge of open source contributors

for regions in the world where multilingualism is the norm and ICT literacy is not well established



OScott

on the web:

<http://www.tuxtrans.org/oscott/>

discussion list:

<http://tech.groups.yahoo.com/group/oscott/>

List of References

- Bowker, L. (2002): Computer-Aided Translation Technology: a practical introduction. Ottawa: Univ. of Ottawa Press.
- Carstensen, Kai-Uwe; Ebert, C. E. Et al. (2010): Computerlinguistik und Sprachtechnologie: eine Einführung, Heidelberg: Spektrum Akad. Verl.
- Díaz Fouces, Oscar; García González, M. (2008): Traducir con software libre. Editorial Comares.
- Dunne, K. (2006): Perspectives on Localization. Amsterdam/Philadelphia: John Benjamins
- Fähndrich, Ursula (2005): Terminology Project Management. In: Terminology, 11:2. 225-260.
- Guillardau, S. (2009): Freie Translation Memory Systeme für die Übersetzungspraxis: Ein kritischer Vergleich. Translationszentrum der Universität Wien
- Le Crosnier, Hervé; Vannini, Laurent (eds) (2012): Net.lang. Towards the multilingual cyberspace. Caen (FR): C&F éditions.
- Picht, H. (2006): Modern Approaches to Terminological Theories and Applications. Bern: Peter Lang
- Quah, C. K. (2006): Translation and Technology. New York, NY [u.a.]: Palgrave Macmillan.