



International Conference on Language Technologies for All (LT4All)

Enabling Linguistic Diversity and Multilingualism Worldwide,

in the framework of the International Year of Indigenous Languages – IYIL 2019

“Everyone should have the possibility to get access to Language Technologies
in his/her native languages, including indigenous languages”

Preamble:

In the context of this document, Language Technologies (LTs) cover from simple keyboarding technologies, spelling/grammar checkers up to speech and speaker recognition, machine translation for text and audio, speech synthesis, spoken dialog, text and document understanding, generation and summarization, sentiment and opinion analysis, answers to questions, information retrieval and knowledge access, sign language processing, etc.

Language Resources cover all types of data sets such as audio/video recordings, with or without human annotations, textual corpora, lexica, etc. in a machine-readable format. It also covers other modalities such as sign languages.

We assume that the needs of each person should be fulfilled and that particular attention should be paid to people with specific needs.

The first 10 statements from the Strategic Document Drafting Committee:

1. All communities, in particular indigenous communities, should have the possibility to get access to all Language Technologies (LTs) that help people break the digital divide, through efficient access to or production of multilingual knowledge, communication, education, and services, in or using their native languages.
2. Native languages should be usable for communication between humans speaking different languages. Cross-lingual LT solutions exist and should be extended to cover all languages, as needed. This will allow a better mutual understanding while also facilitating the access to foreign cultures.

3. Today only 2% of the (more than) 7000 world languages are LT-enabled. Consequently an underlying risk is that languages that are not LT-enabled may never get access to the digital world and those that have some digital presence but are weakly LT-enabled may quickly face digital extinction to the benefit of the few so-called major languages.
4. Communities should have the capability to define their own needs, expectations, and requirements. Today, 20% of the human languages spoken on the planet cover about 99% of the population¹. It is feared that LT development efforts may depend on the market and community size. It is essential to join forces with experienced communities to pay particular attention to the remaining 80% of the languages², independently from any consideration on economic interest or community demographics.
5. Internet access and digital literacy are severe barriers in many geographical areas and have to be seriously addressed, in agreement with the local communities. Deployment of LT applications and services is not hindered by language technology barriers but mainly by the lack of language policies that should lead to collections of needed and sharable language resources and expertise. For instance, having a writing system, which is often the first building block of many technologies, could be perceived as a pre-requisite, but alternative approaches based on spoken information do exist that need to be further developed.
6. Most of the key LTs exist today as state of the art open source packages that could be customized/tuned/ported to all languages, assuming that needed language resources and expertise are available. Such trend should be supported and encouraged. Nevertheless, they still require more research, technology improvement, easy adaptation and portability methodologies, which should be part of a shared research agenda for all languages. The Research community should pay attention to languages that do not have digitized data to undergird participation and decisions about where to invest resources should carefully consider this perspective.
7. Assessment of LT performance for a given language should be systematically conducted to ensure its usability for real services. LTs that are not mature enough and hence not usable may add more confusion than it brings valuable services. It is important to ensure that, for experimental technologies and prototypes that could be released to the public in languages where performance is not yet demonstrated to reach industry benchmark, a clear information is given.

1 https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers

2 <https://www.ethnologue.com/statistics/size>

8. Education is a Human Right. An important trend has developed recently through the availability of online courses (e.g. MOOC) on important domains but for a few languages. LTs can help develop and improve these education processes by allowing access to non-native materials. Such access can build on all modalities such as audio, video, text, sign languages, etc. LTs can also help to develop educational resources in the native language, by extracting them from archives of texts and stories, from oral or written culture.
9. Language preservation is part of the culture preservation and human heritage. LTs allow to curate resources such as audio/video recordings, textual corpora, grammars, and lexica, etc. that are representative of the culture and language use within the community. Such activity allows to enhance preservation and also to conduct language documentation activities for the benefit of future generations as part of their cultural heritage. Documenting a language also requires analyzing the content collected; LTs can accelerate this process, allowing more resources and more languages to be processed.
10. The revitalization of languages, particularly endangered languages, can be promoted by helping indigenous communities have better access to native resources and knowledge. LTs can also directly help revitalize languages; for instance using technologies such as speech synthesis and computer-assisted language learning allow people to remain exposed to the language sounds and to an accurate pronunciation of the language.

Strategic Document Drafting Committee:

- Gilles Adda, LISN-CNRS, France
- Khalid Choukri, ELRA, France
- Joseph Mariani, LISN-CNRS, France
- H  l  ne Mazo, ELRA, France
- Sakti Sakriani, Japan Advanced Institute of Science & Technology (JAIST), Nomi, Ishikawa, Japan