Automatic speech recognition for under-resourced languages: A survey

Laurent Besacier [a], Etienne Barnard [b], Alexey Karpov [c], Tanja Schultz [d]

[a] *Laboratory of Informatics of Grenoble, Grenoble, France*
[b] *North-West University, Vanderbijlpark, South Africa*
[c] *St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russia*
[d] *Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany*

Abstract

Speech processing for under-resourced languages is an active field of research, which has experienced significant progress during the past decade. We propose, in this paper, a survey that focuses on automatic speech recognition (ASR) for these languages. The definition of under-resourced languages and the challenges associated to them are first defined. The main part of the paper is a literature review of the recent (last 8 years) contributions made in ASR for under-resourced languages. Examples of past projects and future trends when dealing with under-resourced languages are also presented. We believe that this paper will be a good starting point for anyone interested to initiate research in (or operational development of) ASR for one or several under-resourced languages. It should be clear, however, that many of the issues and approaches presented here, apply to speech technology in general (text-to-speech synthesis for instance).
© 2013 Published by Elsevier B.V.

Keywords: Under-resourced languages; Automatic speech recognition (ASR); Language portability; Speech and language resources acquisition; Statistical language modeling; Crosslingual acoustic modeling and adaptation; Automatic pronunciation generation; Lexical modeling

1. Introduction

Nowadays, computers are heavily used to communicate via text and speech. Text processing tools, electronic dictionaries, and advanced speech processing systems like text-to-speech (speech generation) and speech-to-text (speech recognition) systems are readily available for several languages. There are however more than 6900 languages in the world and only a small fraction offers the resources required for implementation of Human Language Technologies (HLT). Thus, HLT are mostly concerned with languages for which large resources are available or which have suddenly become of interest because of the economic or political scene. Unfortunately, most languages from developing countries or minorities received only little attention so far. One way of improving this "language divide" is to do more research on the portability of speech and language technologies for multilingual applications, especially *for under-resourced languages.*

This paper is a review on automatic speech recognition (ASR) for under-resourced (UR) languages, which have shown a growing interest in the recent years. While the task of ASR is rather specific, some issues addressed in this paper apply to other HLT tasks as well. This paper is organized as follows: After an *Introduction* that focuses on the language diversity and on our motivation to address the topic, Section 2 gives a brief definition of what we call "under-resourced languages", as well as the challenges associated to them. Section 3 is a literature review of the recent contributions made in ASR for under-resourced languages. Examples of past projects on this topic are given in Section 4, while Section 5 presents the future trends when dealing with under-resourced languages. Finally, Section 6 concludes this work.

1.1. Languages of the world

Counting the number of languages in the world is not a straightforward task. First, one has to define what makes a language, for example to decide if dialects are considered to be a language, if so, which ones should be added, or if not, to draw the line between a language and a dialect. An

estimate for the total number of living languages in the world can be found on the Ethnologue[1] web site. They define a living language as "one that has at least one speaker for whom it is their first language". So, extinct languages and languages that are spoken as a second language are excluded from these counts. Based on this definition, Ethnologue lists 6909 known living languages. This list includes 473 languages that are classified as nearly extinct, i.e. when "only a few elderly speakers are still living". It is important to note that Ethnologue's list includes both verbal and visual-kinetic spoken languages. The latter ones are known as sign languages, which are used for everyday communication by the deaf; these spoken languages combine hand gestures with lips articulation and facial mimics. Almost all countries over the world define their own national sign languages.

Counting how many languages have a written form is also subject to controversy. The foundation for endangered languages web site[2] mentions 2000 written languages by counting published bibles (entirely or portions) but this also includes non-living languages. Omniglot,[3] an online encyclopedia of writing systems and languages, lists less than 1000 written languages and gives details on more than 180 different writing systems.

While counting languages is a tricky task, the number of "well-resourced languages" can be easily given by listing how many languages are identified for core technologies and resources, such as: Google Translate (63 languages involved[4] in 2012), Google search (more than one hundred languages in 2012), Siri ASR application (8 languages in 2012), Wiktionary[5] (∼80 languages in 2012), Google Voice Search (29 languages and accents in 2012).

1.2. Language extinction

In today's globalized world, languages are disappearing at an alarming rate. Crystal (2000) estimated that over the next century about half of all existing languages will be extinct. On average, one could say that every two weeks one language dies. A survey by the Summer Institute of Linguistics (SIL) from February 1999 revealed that about 51 languages are left with only one speaker, 500 languages have 500 speakers left, and 3000 languages have less than 10.000 speakers left. The graph below summarizes the estimates of speakers over languages from the SIL survey. It shows that 96% of the world's languages are spoken by only 4% of its people.

History has shown that not even a language with 100.000 remaining speakers is safe from extinction (Crystal, 2000). The survival of a language depends on the pres-

sure imposed on that language and on its speakers. Pressure may arise from disasters (earthquakes on Papua New Guinea killed several languages), genocide (about 90% of America's natives died within 200 years of European conquering) or simply from the dominance of another language. The latter may result in cultural assimilation (social, political or economic benefits to speak the dominant language) that usually leads to the loss of the suppressed language within few generations (e.g. second generation immigrants).

How could language extinction be slowed down and what are the associated costs. First of all, a language can only be saved if the community itself wants it and the surrounding culture respects this wish. Typically, the community is then supported to fund courses, materials, and teachers. In addition, linguists go into the field, collect and publish language related information such as grammars, dictionaries, speech recordings, and make them available to the public at large. The associated costs depend on the particular conditions, for example if the language has a writing system, etc. Crystal estimates about USD 80.000 per year per language. Considering 3000 endangered languages this would add up to more than USD 700 Million. Organizations like the Foundation of Endangered Languages (FEL) and large-scale UNESCO projects have been established to raise both, attention and funds, to tackle this major challenge (see Fig. 1).

1.3. Good reasons to address less prevalent languages

Some languages might be more attractive than others for Human Language Technologies (HLT). However, for the reasons described above, there are good reasons for developing speech recognition (and other technologies like machine translation) systems for literally all languages in the world. First of all, spoken language is the primary means of human communication. Both, individual and community memories, ideas, major events, practices, and lessons learned are all preserved and transmitted through language. Furthermore, language is not only a communication tool but fundamental to cultural identity and empowerment. So, language diversity in the world is the basis of our rich cultural heritage and diversity. If the world loses a language, the memories and experiences of this culture go with it. Crystal claims that language diversity should be treated like bio-diversity as history has shown that the more diverse eco-systems are strongest.

Human Language Technologies have a lot to offer to revitalize and (at least) document languages and thus prevent or slow down language extinction. The existence of technology may raise interest and make the language attractive again to their native speakers. Moreover, in the perspective of saving some endangered languages (some mostly spoken and not written), the possibility to rapidly develop ASR systems to transcribe them is an important step for their preservation and would facilitate access to audio contents in these languages. A second reason why

---