Language Engineering for Basque in a Visual Communication Technologies Context

Maider Lehr[1], Kutz Arrieta[1], Andoni Arruti[2]

[1]VICOMTech Research Centre,
Mikeletegi Pasealekua 57, 20009, Donostia-San Sebastian
[2]Signal Processing Group, University of the Basque Country,
Manuel de Lardizabal Pasealekua 1, 20018, Donostia-San Sebastian
mlehr@vicomtech.org, karrieta@vicomtech.org, andoni.arruti@ehu.es

Abstract

The integration of language engineering in other applications is gaining support in European research centers and government agencies dedicated to the creation and management of research resources. In this context and given the particular suitability of the Basque Country to understand and promote this type of development and integration, the Basque Government and other institutions are making the necessary efforts and have considered this as one of their most relevant lines of development for the coming years.

In this context, VICOMTech, an applied research center located in Donostia-San Sebastian (Basque Country) has opened a new emerging area in language engineering and intends to integrate in the other areas that the center develops. Therefore, the inclusion of Natural Language Processing devices within applications developed in Digital TV, Multimedia services, Biomedical Sciences, Industrial Applications, and Human-Computer Interaction, will offer added value and contribute to the intelligence of these applications.

This paper is intended to inform the reader about the efforts VICOMTech is making to develop this approach and reports on some of the research already done in the field of speech, in which VICOMTech has already some experience.

Index Terms: language engineering, speech technologies, Basque

1. Introduction

One of the tasks that Language engineering (Computational Linguistics) tries to accomplish is; overcoming the linguistic constraints imposed by the variety of the existing human languages, allowing therefore everyone to use his/her native language to interact with the technology, improving accessibility to the Information Society.

New markets and research interests are opening in the area of Natural Language Processing and Speech. Thanks to the emergence of these new markets this field is making considerable advances. Minority languages cannot afford to stay out of this race, since their survival may depend on this, among other factors. Basque does not count on the resources and devices that have been created for widely spoken languages, but in the last few years Basque has made important advances in the integration to these technologies. Due to the small market for Basque, it is very important, on the one hand, to take actions to raise the demand promoting both the knowledge and the market of existing products. The market must offer in Basque the applications that are available in majority languages such as Spanish, French or English. On the other, we must have a multilingual product to market to the world.

The science, technology and innovation plan for 2007-2010 proposed by the Basque Government makes special mention of Language Engineering. In reference to this, we must mention AnHitz, a strategic project promoted and partially financed by the Basque Government, where VICOMTech [1] is the leader and Elhuyar Fundazioa [2], Robotiker [3], IXA (University of the Basque Country) [4] and Aholab (University of the Basque Country)[5] are partners.

VICOMTech is an applied research center located in the technological park of Miramon in Donostia. It is a non-profit association founded by the INI-GraphicsNet Foundation and Basque Television, Radio and Broadcasting group (EITB) [6].

As mentioned above; one of the recently created strategic research interests in VICOMTech is the promotion of linguistic technologies in multilingual environments, with particular attention to the integration of Basque. AnHitz's goal is the development of linguistic technologies in Basque to enable human-machine interaction, as well as, knowledge management and other applications using this language. The newly created linguistics department is conceived as a transversal technology that should be integrated into and complementing the already existing research fields at VICOMTech.

- Digital TV and Multimedia services: this area specializes in transmission and interactivity standards (DVBT, C,S,H - MHP, etc.), A/V content analysis and management and virtual/augmented reality services for broadcast professionals. The department has some experience in user interfaces for television. Making use of this experience and in order to take advantage of the interactivity offered by the standards used, linguistic technologies will be integrated. First, the know-how in A/V content analysis and management will be complemented with the linguistic tools, developed in order to expand the possibilities for application environments.
- Biomedical applications: this department concentrates on research and development for the healthcare and biotechnology sectors. The main research lines include the most recent advances in image analysis (image processing, segmentation, registration and fusion), visualization (virtual reality and augmented reality), and biomedical information management (transmission, representation, standards and interface). This department is envisioning the integration of linguistic components and

ontologies treatment, along with medical imaging interpretation, to explore Indexation and Retrieval applications for medical and biomedical data.

- Tourism, heritage and creativity: this department designs and implements applications for the creation of interactive digital experiences, providing an added value to the services offered by the tourist, cultural, and creative sectors. Among the key technologies we will mention the following: Virtual and Augmented Reality technologies, mobile applications based on location-based tracking tools for content personalization, semantic-based searching algorithms, content annotation and indexing, and standard based system for multimedia content creation and management. This area has considerable experience in ontology standards treatment and indexing, which has already included linguistic tools for one of its projects with a Basque repository of multimedia patrimonial contents.
- Interaction for education, leisure and e-inclusion: this area develops technologies related to multimodal human - device (PC, PDA, mobile phone or TV) interaction through body and natural language. Here is where speech technologies, face and body animation of the virtual characters, emotional interaction, and natural language processing, are hosted.
- Industrial applications: this department concentrates on industrial applications for VICOMTech's main technologies, such as, interactive 3D computer graphics, Virtual and Augmented Reality, advanced visualization devices, and interactive simulation in environments, for industrial design, manufacturing, and commerce. This area is also, developing an application integrating semantic information and image processing for industrial design, brand, and information monitoring.

2. Speech related research lines

The Speech group in VICOMTech is part of the interfaces research group and Speech is perceived accordingly, as an interface. VICOMTech does not work in the development of synthesizers or recognizers, our goal is to adapt, modify, and/or extend, existing technologies to our context. Given the fact that we work with Basque, we quickly enter the realm of research by having to convert existing technologies to a language with little resources and of a "one of a kind" type of language, quite different from the majority languages.

2.1. Television driven technologies

2.1.1. Subtitling

Subtitling by means of open or closed captions is, for millions of hearing-impaired people, the most useful representation means for speech content in TV and other audio-visual media.

However, most Spanish TV channels provide subtitles as closed captions in teletext format only for some of their prerecorded programs. For live programs, such as sports events and news broadcasts, subtitles are rarely available. Specially trained stenographers and fast typists for live subtitling are expensive.

There are some software solutions for ASR-based live subtitling, e.g. Protitle Live (NINSIGHT) and WinCAPS (Sysmedia) allowing a trained speaker to dictate live subtitles into a trained ASR system ("re-speaking"). Nevertheless, there is no ASR-based system in use for fully automated subtitling.

We studied the feasibility of using dictation for literal transcription of speech for fully automated live subtitling bypassing the re-speaker [7]. To do this, we integrated commercial best-of-its class ASR software with a professional subtitle generator and preprocessing modules we developed . We evaluated the quality of ASR for Spanish, measured the delay between speech and subtitle, and detected particular drawbacks in the components used for subtitling.

Evaluations of the video by means of questionnaires were performed by nine volunteers (aged 25 to 65 years from different social groups) from organizations of deaf and hard of hearing people in Donostia-San Sebastian (Spain).

An objective quality measure of the transcription obtained by the prototype is the word recognition rate, WRR = (H - I) / N, where N is the number of words in the reference, I is the number of the insertions, H is N - (S + D), the number of correctly recognized words, being S the number of substitutions, and D the number of deletions.

Furthermore, we measured the time delay between speech and the appearance of the corresponding subtitle.

We concluded that ASR systems need improvements in:

- Speaker and material independence. It only works well when trained to recognise a single voice (acoustic models) and when trained previously with material related to the contents of the programs (language models and dictionaries).
- Current methods for noise filtering, speech, music and silence detection.
- Speech recognition for multiple speakers.
- Lack of automatic punctuation.

We are, therefore, working on:

- Automatic punctuation. Several algorithms will be developed to segment the audio signal in significant utterances. Three lines will be studied:
  - Detection methods of segmental borders focussing on acoustic features.
  - Detection methods of segmental borders focussing on linguistic features.
  - Detection methods of segmental borders focussing on prosodic features
- Speakers discrimination and identification. Usually, in speaker recognition tasks the speakers to be recognized are known. In other applications, like this one, neither the identity nor the number of speakers are known in advance. The speech signal stream is continuous. Speaker changes must be detected and the stream segmented. This must be done off-line.

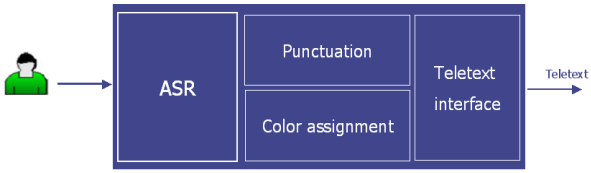

Figure 1: *Subtitling generation system.*

- Automatic generation of captions. We are exploring the possibility of automatically performing all the linguistic

changes that trained subtitlers introduce in the text in order to make it really understandable for hearing-impared viewers.

2.1.2. *Voice Transformation*

Voice Transformation is relevant for a wide range of speech related applications such as:

- Speech processing
  - TTS systems
  - Restoration of old recordings
  - Training of speech recognition systems
  - Speaker verification/ identification systems
- Multimedia/ music
  - Karaoke
  - Voices for virtual environments/ chats
- Dubbing and looping
  - Preservation of the original voice of the actor/ actress
  - Increase of voice registry
  - Looping

Dubbing companies need experienced and qualified people, with some specific features (age, gender, voice depth). Dubbing requires specific skills, with specific registries, a specific voice quality, as well as a good dramatization abilities. So, difficulties to get new voices are considerable. For example, often, adult voices, mostly female adult voices, are used to dub children voices. Dubbing companies must overcome this challenge: to find solutions for this lack of available voice registries. This voice and registry limitations are more noticeable in minority languages.

In this context we proposed to develop a prototype with a friendly interface. The interface provides the possibility of changing a source voice to a target voice in an intuitive and friendly way. We are working with some Basque dubbing companies which work closely with EITB (Basque Television) to create a system capable of generating different voice registries for dubbing in TV and cinema [8].

2.1.3. *Synchronization of audio and animation*

In human-machine interfaces there is a clear trend to merge different possibilities of presenting information, in particular, speech and facial animation. The presence of minority languages in the area of virtual characters is very limited. This is obviously due to the lack of both resources and tailored technology. We have developed a system capable to produce suitable data for the animation of faces from natural voice in Basque using open source technologies [9]. The output of the speech analysis performed matched a set of visemes (visual representation of the phoneme) and phonetic data, corresponding to the lip visualization of the virtual character for each frame of the animation. This output was used to synchronize the animation with on-line audio in real time. The application captures the speech signal from the input through a sound card and identifies the appropriate phonemes. As phonemes are recognized, they are mapped to their corresponding visemes. The virtual character is then animated in real time and synchronized with the speaker's voice. The application consists of three modules Fig. 2:

- The phoneme recognition system.
- The module that sends the input audio to the recognition system.
- The communication interface between the recognition system and the animation platform.

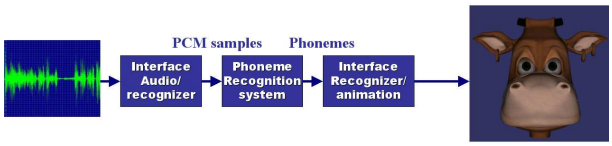

Figure 2: *Facial animation for live speech input in Basque.*

The goal was to obtain a useful and usable application in the television domain, where virtual presenters are more and more common. A quiz type TV program for children using a virtual character was created. This character is presently running in a popular Basque TV (EiTB) program, in which children answer questions and interact with the said character. In this case, a cow. Results are satisfactory for this first version [8].

The animation reacts to the caller's answers, it therefore, needs to run in real time. Lip animation runs as the actress who dubs the virtual character in the program speaks into the microphone. This voice is sent in real time to the software developed in this project. The software analyses the stream and generates the data to synchronize the lips with the audio. This information is interpreted by the animation engine.

2.2. Other applications

2.2.1. *Multimedia Information Retrieval*

The very essence of an information retrieval (IR) system is to satisfy the user's information needs, expressed by a query submitted to the system. There is now widespread use of information retrieval (IR) techniques to access information stored in electronic format. One of the most widely used examples of this is in internet search engines.

There is also a great deal of information in audio format. One such area is the audio associated with radio and television news broadcasts. If these audio sources could be transcribed automatically then the information they contain can be indexed and relevant portions of broadcasts retrieved using conventional IR techniques. In collaboration with EITB, the Basque radio, television and broadcasting group, we are working in a project which consists on developing an information retrieval system to efficiently access multimedia contents.

The approach is to combine speech and more "traditional" IR techniques to get to the desired results, in order to manage and use large catalogs of multilingual multimedia contents. Among other techniques (image analysis, etc.) we are going to use speech recognition systems to convert the spoken word to a text transcription which is then passed on to a regular text-based IR search engine. Speech recognition technologies in Basque are far from the existing technologies for other languages and we are not in the business of developing such technologies. We need to adapt to the situation and, therefore we will proceed in different scenarios, integrating separate applications:

- CLIR: using existing partial Machine Translation applications for Basque we will preserve the ability of the user to perform the query in Basque and receive:

  - Results in Basque and other languages when dealing with parallel text.
  - Results in Basque and other languages (the ones in Basque being of poorer quality) when not dealing with parallel text.
  - Results in languages other than Basque

  In principle, these other languages are Spanish, French and English.

- IR for Basque combining several sources: sound, image recognition and text.
- IR for Basque accessing the Internet or only the Basque Science and Technology repository using speech only as an interface.

We also plan to include some type of mutual feeding of the text, image and audio contents of these catalogs.

2.2.2. Autopuntuation

As mentioned before, we are exploring the possibility of generating sentence punctuation automatically combining speech and linguistic technologies. This is quite an ambitious project and our wish would be that, for once, Basque would the first language having an application that other languages lack.

But this is not our only motivation, it is clear that such a development can, in the one hand, be extended to other languages, with promising commercial consequences, and generate knowledge on the linguistic and acoustic aspects of Basque and this type of application in general, in the other.

3. Conclusions

More and more users want and need to interact with devices in more natural and easy ways. Also, the amount of catalogs and knowledge databases has increased. This information is also stored in different formats and languages. This information needs to be stored, managed, retrieved and understood. Natural language Processing is an unavoidable component of the applications of the future. This field has gained great relevance in European research programs. The Basque Government, research centers, universities and public and private companies are showing interest in Speech and Natural Language Processing related applications.

We have tried to present here some of the projects we have worked on, mostly in Speech, and our present and future projects geared towards combining "forces": Speech, NLP, Image Analysis, Virtual Characters, Culture, Ambient Intelligence, etc., not only to promote social and industrial progress, but also to reduce the technological gap existing between Basque and other languages.

4. Acknowledgements

The projects mentioned in this paper have been partially supported by the Basque Government through the ETORTEK and INTEK programs of the Department of Industry and by the Spanish Ministry of Industry, Tourism and Commerce.

Furthermore, we would like to thank some our partners and clients such as EITB, Irusoin, Mixer, Baleuko, Talape, Elhuyar, Robotiker, Aholab, IXA, Antena 3 TV, AudioText and Ceapat.

5. References

[1] http://www.vicomtech.org/.

[2] http://www.elhuyar.org/.

[3] http://www.robotiker.com/.

[4] http://ixa.si.ehu.es/Ixa.

[5] http://aholab.ehu.es/aholab/Home/.

[6] http://www.eitb.com/.

[7] Obach, M., Lehr, M., Arruti, A., "Automatic Speech Recognition for Live TV Subtitling for Hearing-Impaired People", 9th European Conference for the Advancement of Assistive Technology in Europe (AAATE 2007), Assistive Technologies Research Series Volume 20, 2007, pp. 286-291.

[8] http://www.vicomtech.es/castellano/html/videos_demos/index.html.

[9] Lehr, M., Arruti, A., Ortiz A., Oyarzun D. and Obach, M., "Speech Driven Facial Animation using HMMs in Basque", Text, Speech and Dialogue, Proceedings Lecture Notes in Artificial Intelligence (Springer), 2006, pp. 415-422.