AidSLT : A Spoken Language Translation Tool for Data Collection in Humanitarian Settings

ARMANDO, Alejandro

Abstract

This master’s dissertation is concerned with the development and evaluation of AidSLT, a spoken language translation tool designed to collect data in humanitarian settings. The central idea stems directly from and draws upon the work done by the University of Geneva on MedSLT (Bouillon et al., 2005), an analogous tool intended for use in doctor-patient diagnosis dialogues. The author describes a first prototype of the tool designed for administering multilingual household surveys about malaria prevention measures and present the results of a quantitative and qualitative evaluation for the language pair English (L1) French (L2).

Reference


Available at:
http://archive-ouverte.unige.ch/unige:75455

Disclaimer: layout of this document may differ from the published version.
AidSLT: A Spoken Language Translation Tool for Data Collection in Humanitarian Settings

Alejandro ARMANDO

Master of Arts in Translation
Director: Pierrette Bouillon
Jury: Emmanuel Rayner
Acknowledgements

I first want to thank Prof. Pierrette Bouillon, whose fascinating course on linguistic engineering made me change the area of specialization of my master’s program and begin this research project. Her encouragement, patience and support have been invaluable during these past two years. I also want to thank Dr. Emmanuel Rayner and Dr. Nikolaos Tsourakis, who shared their vast experience with me and made me feel part of their team.

I am also indebted to all the humanitarian workers who tested the prototype of the tool and answered the questionnaire. In particular, I would like to acknowledge the contributions of Ivan Vuarambon, from the United Nations High Commissioner for Refugees, who showed interest in my project from the beginning and provided precious feedback.

Finally, to my family and friends, I could not have finished this project without your support and understanding. I am lucky to have you.
## Contents

Introduction .................................................................................................................. 1

I. Automatic Speech Recognition ............................................................................... 3

1.1 A brief history of speech recognition ................................................................ 4
1.2 Main processing phases ...................................................................................... 5
  1.2.1 Signal Processing ......................................................................................... 6
  1.2.2 Acoustic models .......................................................................................... 8
  1.2.3 Language models ....................................................................................... 16

1.3 Conclusion ........................................................................................................... 22

II. Machine Translation ............................................................................................... 23

  2.1 Translation problems ....................................................................................... 23
  2.2 Main approaches to MT ................................................................................... 24
  2.3 SLT Systems .................................................................................................... 28

  2.4 Conclusion ....................................................................................................... 30

III. AidSLT ..................................................................................................................... 31

  3.1 Motivation ....................................................................................................... 31
  3.2 Context of use .................................................................................................. 32
  3.3 Corpus .............................................................................................................. 34
  3.4 Architecture of the system .............................................................................. 37

  3.5 User interface .................................................................................................. 42

  3.6 Conclusion ....................................................................................................... 45

IV. Evaluation ............................................................................................................... 46

  4.1 Methodological aspects ................................................................................... 46
  4.2 Objectives ....................................................................................................... 51

  4.3 Design .............................................................................................................. 52

  4.3.1 Pilot testing ............................................................................................... 53

  4.3.2 Collection Method ..................................................................................... 54
List of figures

Figure 1. Radio Rex.......................................................... 4
Figure 2. Milestones in Speech Recognition.............................................. 5
Figure 3. The speech recognition process (adapted from Rayner et al., 2006, p. 6) .......... 6
Figure 4. Continuous and discrete signals (adapted from Johnson, 2011)............................ 7
Figure 5. Acoustic Modelling.................................................................. 9
Figure 6. A waveform and a spectrogram.................................................. 11
Figure 7. Pronunciation dictionary.......................................................... 14
Figure 8. Variation and homophony......................................................... 16
Figure 9. Google Ngram Viewer.................................................................. 19
Figure 10. Linguistic MT systems (Hutchins & Somers, 1992)......................... 27
Figure 11. The Phraselator and MediBabble............................................... 29
Figure 12. The MICS survey..................................................................... 36
Figure 13. The MIS survey...................................................................... 36
Figure 14. Group 7. WaterForDrinking....................................................... 38
Figure 15. Extract from instructions. WaterForDrinking............................... 39
Figure 16. The Next field......................................................................... 39
Figure 17. QuestionTemplate. WaterForDrinking........................................ 40
Figure 18. Template applications............................................................. 41
Figure 19. Answers................................................................................ 42
Figure 20. USE THE APPLICATION tab.................................................... 43
Figure 21. PROVIDE ANSWER tab........................................................... 44
Figure 22. Sampling by sentence length.................................................... 56
Figure 23. Sampling by type of sentence................................................... 56
Figure 24. Sampling by subject............................................................... 57
Figure 25. Pop-up message....................................................................... 59
Figure 26. Freely expressed interactions.................................................... 64
Figure 27. Out-of-coverage input.............................................................. 65
List of tables

Table 1. A minimal conceptualization of the initial acoustic-level phases of the speech recognition process.................................................................................................................................................... 14
Table 2. Linguistic ambiguity........................................................................................................................................................................................................... 17
Table 3. Context free rules. .............................................................................................................................................................................................................................. 21
Table 4. Vector Control Indicators (World Health Organization et al., 2013, p. 7) ................ 34
Table 5. Objectives, attributes and metrics............................................................................................................................ 52
Table 6. Sample. .............................................................................................................................................................................................................. 58
Introduction

Translating spoken language is perhaps the single most ambitious endeavor that can be undertaken in the field of Natural Language Processing, so much so that as recently as twenty years ago it was considered simply too difficult for all but the most basic research applications (Arnold, 1994, p. 8).

The development of a functional Spoken Language Translation (SLT) system involves addressing all fundamental problems associated to each of its individual component technologies—speech recognition, machine translation and speech synthesis—as well as all additional problems that result from integrating such technologies in an efficient way.

Recent computational advances, particularly in the field of large-vocabulary speech recognition, have allowed the research community to envision a much wider range of potential applications for SLT. In its brief history, the field has gone from very rudimentary speaker-dependent systems that targeted highly restricted domains, to aiming at open-domain multilingual conversations between Skype users and obtaining allegedly impressive results¹.

The technological progress made in recent years is phenomenal and it certainly feels like an exciting time for speech translation. However, current systems are still far from being perfect. Even the most performing applications yield a percentage of error that makes them inadequate for safety-critical domains where language barriers pose a real problem. One example is translation in medical contexts. Another one is the intended context of use of the tool we present here.

This master’s dissertation is concerned with the development and evaluation of AidSLT, a spoken language translator designed to collect data in humanitarian settings. The central idea stems directly from and draws upon the work done by the University of Geneva on MedSLT (Bouillon et al., 2005), an analogous tool intended for use in doctor-patient diagnosis dialogues.

We describe a first prototype of the tool designed for administering multilingual household surveys about malaria prevention measures and present the results of a quantitative and

¹ blogs.microsoft.com/blog/2014/05/27/microsoft-demos-breakthrough-in-real-time-translated-conversations (accessed in August 2015)
qualitative evaluation for the language pair English \( (L_1) \) French \( (L_2) \), conducted by a small group of humanitarian workers between March and July 2015.

The main purpose of the evaluation was to determine the viability of the project, i.e., if the tool would be well received by potential users and whether it was likely to be deployed with some degree of success in its intended context of use.

The structure of this dissertation is as follows. Chapter I presents a thumbnail history of Automatic Speech Recognition (ASR) and explores its main processing phases, namely: Digital Signal Processing, Acoustic Modelling and Language Modelling. It concludes by outlining the ASR strategy deployed by AidSLT. Chapter II presents an overview of the main problems and approaches involved in Machine Translation (MT), and presents a few considerations about integrating this technology in a speech translation system. It concludes by describing the translation strategy in AidSLT. Chapter III describes the first prototype of AidSLT. It includes information about the factors that motivated this study, the criteria that were used to build the corpus, the architecture of the system and the user interface. Chapter IV describes several aspects related to the design of the evaluation. Chapter V presents the results and Chapter VI presents general conclusions provides an outlook for future research.

Part of the material presented in Chapter III has been already published in the proceedings of the 36th session of the ASLIB Conference (Armando et al., 2014).
Chapter I

I. Automatic Speech Recognition

The overall goal of this chapter is to lay the groundwork for discussing the most salient architectural features of AidSLT regarding speech recognition, which will be briefly introduced at the end of this chapter and fully dealt with in Chapter III.

This chapter is organized as follows. Section 1.1 presents a brief history of speech recognition that highlights the dramatic evolution of this technology and its enormous contribution to the field of speech translation. Section 1.2 describes its main processing phases. Section 1.2.1 and Section 1.2.2 introduce a number of basic notions about Digital Signal Processing and Acoustic Modelling that are intended to provide readers with an intuitive global grasp of the initial acoustic-level phases of the speech recognition process. The author has included a considerable amount of information about these topics because he has found that they are often absent in the literature on computational linguistics or presented in a way that comes off as too complicated to follow. Section 1.2.3 provides a complete account of the language-modelling task. Finally, Section 1.3 concludes by presenting a brief characterization of the ASR strategy deployed in AidSLT.
1.1 A brief history of speech recognition

The first commercialized speech recognition system was launched to the market in 1972 by an American company called Threshold Technology Inc. at a price of USD 20,000. This system was called VIP100 and was designed for voice command applications. It could recognize “a hundred isolated words after a long and tedious training phase and occupied a volume of approximately one cubic meter” (Mariani, 2010).

Roughly fifty years before that, the appearance of a children’s toy called Radio Rex marked the first time in history that speech recognition technology was applied to a commercial product. The toy was a dog made of celluloid that would leap out of its house at the sound of its name. An electromagnetic circuit resonant to 500Hz —the frequency of the vowel /e/—made the dog exit the house (Jurafsky & Martin, 2009).

The enormous progress made during the past few decades has paved the way for the development of high-quality ASR systems. One major breakthrough in speech technology, which dates from the 1970s, was the development of mathematical models called Hidden Markov Models (HMMs) to determine the probability distribution of speech sounds. A more recent technique developed in recent years, called Deep Neural Networks (DNNs), improved recognition rates by over 30% (Seide et al., 2011).

Current next-generation speech recognizers yield impressive results. For example, the low-end version of Dragon Naturally Speaking, Nuance’s flagship speech recognition software, costs less than 80 dollars and offers an overall accuracy of 99 percent. Today people use ASR systems to operate car navigators, household appliances, cellphones and personal computers; to improve their knowledge of a second language; to overcome speech impediments; or to dictate and process text.
Figure 2 below covers the entire history of speech technology divided into two stretches of about fifty years each.

**Figure 2. Milestones in Speech Recognition.**

### 1.2 Main processing phases

The task of speech recognition consists of converting a speech signal into a string of written words.

State-of-the-art recognizers rely on complex search algorithms that derive information from three distinct sources of knowledge. First, there is *signal processing*, which converts sound waves into digital representations; next, there is *acoustic modelling*, which specifies how utterances in a language are realized as digital representations of sound waves; and, finally, there is *language modelling*, which specifies which words the application is expected to recognize and in what orders they are allowed to occur (Rayner et al., 2006).

These three main processing phases will be dealt with individually and successively in the following three sections of this chapter. Figure 3 below illustrates how they relate to one another.
1.2.1 Signal Processing

The input to a speech recognizer, like the input to the human ear, is a complex series of changes in air pressure caused by the specific way that air passes through the glottis and out of the oral and nasal cavities of a speaker (Jurafsky & Martin, 2009).

These sound-producing pressure fluctuations are first captured by a transducer—a microphone in front of the speaker’s mouth, as illustrated in Figure 3—and then plotted over time into a more or less detailed visual representation that displays the acoustic properties of the speech signal.

The instrumentation used to do this—to record and analyze speech—is computers and computer programs. So when we talk about signal processing, at least in the context of speech recognition applications, we are actually talking about digital signal processing or “the way in which computers handle acoustic signals” (Johnson, 2011, p.49).

With this in mind, the first notion that must be considered is the dichotomy between continuous (analog) and discrete (digital) signals. The differences between these two types of signals can be explained in relation to 1) the way time and amplitude are represented in a waveform and 2) the different types of instruments that are used for processing signals.

Figure 3. The speech recognition process (adapted from Rayner et al., 2006, p. 6).
First, in a continuous signal, time and amplitude are instantiated by numbers that have a theoretically infinite number of places after the decimal point, while in a discrete signal, the number of decimal places is always finite. Thus, the former is best represented by a continuous line having an amplitude value at all points in time, whereas the latter is most accurately represented by a juxtaposition of discrete points, given that it actually constitutes a sequence of amplitude values (Johnson, 2011).

The chart below illustrates such distinction.

![Continuous and discrete sine waves](image)

*Figure 4. Continuous and discrete signals (adapted from Johnson, 2011).*

Second, the terms “continuous” and “discrete” can be seen as synonyms of “analog” and “digital”, respectively, if we think of signals in connection with the taxonomy applied to the instruments that are used to process them. This is neatly explained by Keith Johnson in *Acoustic and Auditory Phonetics:*

Many of the instruments that we use in recording and analyzing speech can convert continuous air pressure variations into continuous electrical signals, but computers can’t do this. Computers can’t store numbers with an infinite number of places after the decimal point, so all acoustic signals in computers are discrete, even though the pressure fluctuations that produce the sensation of sound are always continuous. Instruments like tape recorders that can store a continuous signal are called **analog devices**, because they can store electric analogs of the acoustic signal, whereas instruments that have to convert the continuous signal into a discrete signal are called **digital devices**, because signals get stored as digits (Johnson, 2011, p. 50).
The process by which an analog sound wave becomes a discrete signal that can be stored and manipulated by a computer is referred to as analog-to-digital conversion. This process involves a first step of sampling, which consists of chopping up the speech signal into a series of discrete samples, and a second step of quantization, which consists of measuring the amplitude of the signal at each of those points in time (Johnson, 2011).

Keith Johnson compares this process to manually converting the graph of a sine wave, like the one shown in Figure 4, into a table of numbers using a ruler to take measurements of the wave form.

First, you mark off equal intervals on the time axis, and then at each tick on the time axis you take a measurement of the amplitude of the wave. Marking intervals on the time axis corresponds to sampling […] and measuring the amplitude of the waveform at each time point corresponds to quantization […]. (Johnson, 2011, p. 51)

Sampling and quantization are important phases of the speech recognition process because they allow us to capture the physical properties that are needed to study the signal. However, these will not be discussed in detail because the main issues involved, such as the sampling rate or the accuracy of the amplitude measurements, are well beyond the scope of this dissertation.

Two basic notions that the reader should note are: 1) that a speech signal is stored in a computer as a series of digits that contain all the auditorily important information; and 2) that this numeric version can then be used to generate various kinds of visual representations that display the acoustic properties of the signal.

The best known visual representation of sound is most probably the waveform. However, many computational applications, including automatic speech recognition, rely on more complex spectral plots that display the frequency components of the sound wave. Spectrograms are one example of such plots; they are the most effective and widely-used tool for the acoustic analysis of speech and will be considered in more detail later on in this chapter.

1.2.2 Acoustic models

Acoustic models define the relationship between the structure of an acoustic signal and the distinct sounds of a language (Chen & Jokinen, 2010, Chapter 12) and associate sequences of
speech sounds with individual words, such as the sequence /ˈf l aʊər/ and the word *flower* in the example below.

*Figure 5. Acoustic Modelling.*

The first phase of the acoustic model of a speech recognizer, delimited by a red line in Figure 5, constitutes the **acoustic modelling** task in itself, i.e., the process by which phonemes can be identified in the spectral representation of an acoustic signal; this process can also be referred to as phoneme identification. The second phase consists in building a **pronunciation dictionary** or **lexicon** that associates sequences of phonemes to individual words. Both phases will be dealt with successively in the present section.

An understanding of the first phase of the process requires some consideration of how linguistic sounds are produced by the articulators of the human vocal tract and how they are realized acoustically.

All sound waves are produced by the vibration of a source and intensified or damped by the body around them. In a trumpet or a flute, for example, sound waves are produced by the vibration of air at the mouthpiece and then amplified as they pass through the body of the instrument. Likewise, speech sounds are produced by the vibration of a source—typically the vocal folds—and then amplified by the resonating cavities in the vocal tract, namely: the pharynx, the mouth and the nose.

What causes certain components of the speech sound to be amplified, or damped, is the particular shape of the resonating cavities of the speaker, which in its turn is determined by the position of the tongue and other articulators in the vocal tract.

In the case of vowels, the parameters that determine the shape of the vocal tract are **tongue height**, which indicates how close the tongue is to the roof of the mouth; **tongue frontness** or **backness**, which indicates whether the tongue is positioned towards the front or the back of the
mouth; and lip protrusion, which indicates whether the shape of the lips is rounded or not at the moment of producing the vowel sound (Johnson, 2011).

The correlation between the shape of the vocal tract and the quality of vowel sounds is made evident by the fact that it is simply impossible for a speaker to pronounce an /ae/ sound having set their vocal tract to the inherent configuration of an /i/ sound (or any other speech sound for that matter; naturally, the same principle would apply to all analogous combinations of any given speech sound and any divergent or “wrong” configuration of the vocal tract).

In the specific case of /ae/ and /i/, if one would care to entertain this idea for a second longer, it is by drawing the tongue closer or less close to the roof of the mouth that a speaker gets to modify the size and shape of the oral cavity and, thus, determine the quality of the resulting vowel sound. The /i/ sound is produced with the tongue placed near the palate, while the /ae/ sound is produced with the tongue in a much lower position.

This account of speech production, i.e., conceiving of the vocal tract as a natural acoustic filter that modifies the sound made by the vocal folds, is known as the source-filter theory or source-filter model of speech production.

The central idea is that filters block or let pass components of sound of different frequencies. As the sound wave travels from the vocal folds to the exterior of the mouth, it “plucks” other sound waves from the resonating cavities in the vocal tract, forming a bundle of frequency components. The frequency value produced at the source is called the fundamental frequency (FO), and the frequency values produced at the resonating cavities are called formants or harmonics.

Frequency components are essential to automatic speech recognition — as they are to the perception of speech by humans — because they account for a large portion of the phonetic quality or the “identity” of the resulting speech sound, and because they can be made visible in various spectral representations.

Figure 6 shows a waveform view and a spectrogram of two sequences of English speech sounds separated by a short pause. The sounds were pronounced by the author and recorded with specialized software for phonetic analysis.

---

To round off the brief information provided in the previous section, a spectrogram is “a way of envisioning how the different frequencies that make up a waveform change over time” (Jurafsky & Martin, 2009, p.262). The horizontal axis shows time, as in the case of the waveform, and the vertical axis shows frequency in hertz. Amplitude is represented by the intensity of each point in the image, varying from black at the strongest intensity (the highest amplitude value) to white at the weakest (the lowest amplitude value).

As can be observed, the phoneme /i/ presents a constant pattern in all four instances: a low first formant (F₁) occurring at about 260 Hz, and a very high second formant (F₂) occurring at about 2450 Hz. The location of these two frequency components — represented by visibly darker bars marked with green lines — can be said to be the characteristic spectral pattern of the /i/ sound and, consequently, can be used to distinguish it from other vowel sounds.

In fact, all vowel sounds can be recognized in a spectrogram by analyzing the disposition of their first two formants.

Consonants, for their part, are different in that the layout of formants in a spectrogram is not the sole requirement for their identification. Other acoustic cues come into play as a result of the fact that they involve some kind of restriction to the flow of air coming from the lungs. Before analyzing the spectral representation of /f/ and /v/, we will introduce some preliminary
notions about the production of consonants (just as was done in the case of vowels a few pages above).

From an articulatory perspective, consonants can be classified according to three main criteria: voicing, i.e., whether the vocal folds vibrate or not when the sound is produced; place of articulation, i.e., the specific anatomical point in the vocal tract where the maximum restriction of airflow is produced; and manner of articulation, i.e., how the obstruction of airflow is produced.

Let us begin with /f/. If a speaker places their lower lip loosely on the lower edge of their upper teeth and forces air out of their mouth, they create a partial obstruction to the flow of air coming from the lungs. The noise made by the air escaping past that obstruction will sound exactly like the /f/ sound at the beginning of the words flower and philosophy, or at the end of the word laugh (please note that one single speech sound corresponds to multiple orthographic realizations and keep this in mind when we discuss the notion of ambiguity in the next section).

Since the articulators at play in this process are the lower lip and the upper teeth, /f/ is said to be a labiodental sound. Similarly, because the hissy noise made by the air escaping the lips is caused by friction, the sound is characterized as a fricative. Lastly, /f/ can be characterized as a devoiced or voiceless linguistic sound because it does not require vibration of the vocal folds.

The /v/ sound differs in that it does require vocal fold vibration, but otherwise it behaves exactly as /f/ regarding articulatory properties. Thus, /f/ can be said to be a voiceless labiodental fricative, whereas /v/ can be said to be a voiced labiodental fricative.

This distinction is visible in the spectrogram displayed in Figure 6. The darker bar at the bottom of the plot in the portion that corresponds to /v/, slightly clearer than the other bars, means that the vocal folds were vibrating when the sound was produced, and that such sound was subsequently amplified by the vocal tract of the speaker in a particular way (unsurprisingly, in a manner similar to the way neighboring /i/ was amplified).

Conversely, the portion of the spectrogram that corresponds to /f/ shows no bars. This is because there was no vocal fold vibration involved in the production of /f/ and no vocal tract filtering. The source of the sound in the case of /f/ is to be found in the lips (very far forward in the mouth) rather than in the vocal folds of the speaker.
Another acoustic cue that can help distinguish these sounds from one another is their amplitude. The portion of the spectrogram that corresponds to /f/ is visibly darker than the portion that corresponds to /v/, suggesting that overall amplitude is higher. This difference is perhaps easier to spot in the waveform view, where the chunk of the wave that corresponds to /f/ is bigger than that of /v/.

The articulatory principles underlying this last distinction are related to one of the two main factors that determine the production of turbulent noise, namely: “the volume velocity of the airflow (volume of air going past a certain point per unit of time)” (Johnson, 2011, p. 152), the other factor being the size of the channel that lets the air through. The idea is that the amplitude of turbulent noise, characteristic of all fricative sounds, is determined by the velocity of the air molecules as they pass through a channel (the faster they move, the louder the sound).

Keith Johnson explains the intrinsic differences in amplitude between /f/ and /v/ in terms of their articulatory properties by suggesting a quite revealing practical demonstration.

[...] if you drape a sheet of paper over your head, you can easily blow it off with a voiceless labio-dental fricative [f] but not with a voiced labio-dental fricative [v]. This is because during voicing the vocal cords are shut (or nearly so) as much as they are open. Therefore, given a comparable amount of air pressure produced by the lungs (the subglottal pressure), the volume velocity during voicing is much lower than it is when the glottis is held open (Johnson, 2011, p. 156).

It is interesting to note that volume velocity, the physical property determining the amplitude of /f/ and /v/, is directly related to voicing, the articulatory phenomenon that typically elicits the formation of frequency components. This reveals the complex interplay that exists between the articulation of speech sounds, their acoustic realizations and their visual representations; and makes it clear that a grasp of such relations is key to understanding how acoustic models work.

To further such an understanding, we present three basic notions that offer a summary account of the present section: 1) that spectrograms allow for the recognition of phonemes; 2) that the recognition task is performed on the basis of the intrinsic acoustic properties of speech sounds; and 3) that acoustic properties bear a relationship with articulatory principles.
These three basic notions are meant to complement the two notions about digital signal processing presented at the end of Section 1.2.1 and, all together, provide the reader with an intuitive global grasp of the initial acoustic-level phases of the recognition process.

Table 1 below gathers all five notions.

<table>
<thead>
<tr>
<th>Digital signal processing</th>
<th>A speech signal is stored in a computer as a series of digits that contain all the information that is important for human audition.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This numeric version of the signal can be plotted over time into various kinds of visual representations.</td>
</tr>
<tr>
<td>Acoustic modelling</td>
<td>Some of these representations (spectral ones) allow for the recognition of phonemes.</td>
</tr>
<tr>
<td></td>
<td>The recognition task is performed on the basis of the intrinsic acoustic properties of speech sounds.</td>
</tr>
<tr>
<td></td>
<td>The acoustic properties of speech sounds bear a relationship with articulatory principles.</td>
</tr>
</tbody>
</table>

Table 1. A minimal conceptualization of the initial acoustic-level phases of the speech recognition process.

Finally, the second phase of the acoustic model of a speech recognizer, delimited by a green line in Figure 5 and replicated in Figure 7 below, consists in associating sequences of phonemes with individual words of a language (Rayner et al., 2006).

Figure 7. Pronunciation dictionary.
Building a good pronunciation dictionary constitutes a major task for various reasons. Perhaps the most salient ones are 1) that the number of words in natural languages is very high and 2) that the number of possible pronunciations for those words is even higher.

The challenges associated with the first factor are fairly straightforward; they stem from the sheer size of the undertaking. Pronunciation dictionaries for large-vocabulary applications typically contain hundreds of thousands of entries. Creating such a large number of entries demands great effort and close attention to questions of storage, access and maintenance.

The second reason is explained by the fact that speakers of the same language pronounce certain words differently depending on various factors, such as the geographical region where they live, their socioeconomic status and their age, among many others. These differences in pronunciation are instances of a broader linguistic phenomenon called variation, which is also concerned with the lexical and structural choices of speakers and the factors that motivate them.

Let us consider a specific example. If we were to build a speech translation tool for the catering industry that includes Spanish as an input language—not an unlikely scenario, since many commercial systems target this domain at present—the pronunciation dictionary would have to include an entry for the Spanish word “sándwich” (a loanword from English), along with all of its many pronunciation variants, including the more refined /ˈsandwiʃ/ and /ˈsán gwi tʃ/, but also the less prestigious /ˈsán gwi tʃə/, /ˈsán gu tʃə/ and /ˈsam bu tʃə/. The ratio is seven sequences of phonemes to one individual word.

Since acoustic models are trained from large quantities of recorded and transcribed speech, the success of a speech application in recognizing the various pronunciations of the Spanish word “sándwich” will depend on whether the training data include examples of people speaking those variants.

Some applications, in particular large-vocabulary dictation tools like Dragon NaturallySpeaking, can be trained to recognize the voice of a single user taking into account their idiosyncratic preferences in pronunciation. This typically improves word accuracy, but the complications that result from creating and appropriately activating individual user profiles mean that this method is often hard to implement (Bouillon, Routledge).

Now, the fact that one single word can be associated to several different sequences of phonemes is just one part of the problem. The other part is that one sequence of phonemes (or more) can
be associated to several different words. For example, in French, the sequences of phonemes /vɛ̃/ and /vain/ can both be associated to the words vin (wine), vain (vane), vingt (twenty), vaincs (to defeat, 1\(^{st}\) and 2\(^{nd}\) person singular, present tense), vainc (to defeat, 3\(^{rd}\) person singular, present tense), vints (to come, 1\(^{st}\) and 2\(^{nd}\) person singular, simple past) and vint (to come, 3\(^{rd}\) person singular, simple past). The ratio in this example is two sequences of phonemes to eight individual words.

This phenomenon is called homophony and will be dealt with more carefully in Section 1.2.3, since it is the language model of a speech recognizer that distinguishes acceptable from unacceptable sequences of words.

Figure 8 below picks up the word used as an example in previous figures (3, 5 and 7) to illustrate how multiple sequences of phonemes can be associated with multiple words within a single entry of the pronunciation lexicon. The phenomenon to the left of the arrow concerns the acoustic modelling phase and constitutes an instance of linguistic variation, whereas the phenomenon to the right concerns the language modelling phase and can be described as homophony.

![Figure 8. Variation and homophony.](image)

1.2.3 Language models

Language models specify which words can be recognized by the system and in which order they are allowed to occur. Their goal is to constrain the range of possible utterances that the ASR component is expected to consider so as to improve the overall performance of the recognition engine.

Underlying this modelling phase is a phenomenon that is intrinsic to all natural languages: ambiguity, i.e., the fact that a single linguistic unit, spoken or written, can have multiple interpretations at one or more levels of linguistic description, such as morphology, phonetics, phonology, syntax, semantics, etc. The word ambiguity is used throughout this dissertation as an umbrella term to encompass all instances of such one-to many relationships.
Jurafsky and Martin (2009, p. 4) illustrate the vast heterogeneity of this phenomenon by pointing at the numerous realizations that can be assigned to the spoken utterance “I made her duck”. Table 2 below gathers a number of the interpretations that can be established for this utterance at the lexical, syntactic and semantic levels.

<table>
<thead>
<tr>
<th>Spoken Input</th>
<th><strong>LINGUISTIC AMBIGUITY</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lexical</td>
<td>Morphosyntactic</td>
</tr>
<tr>
<td>/ at med ər dæk /</td>
<td>I made her duck</td>
<td>I cooked the duck she owns.</td>
</tr>
<tr>
<td></td>
<td>I maid her duck</td>
<td>I created the (plaster/clay) duck she owns.</td>
</tr>
<tr>
<td></td>
<td>I’m aid her duck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I’m aid or duck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I’m eight her duck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I’m eight or duck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I’m ate her duck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I mate her duck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eye mate or duck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eye maid her duck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Linguistic ambiguity.**

Homophony is subsumed within a broader phenomenon referred to as lexical ambiguity (Hutchins & Somers, 1992). The second column in Table 2 shows some of the many possible written sentences that can be associated to a single spoken utterance containing several homophones, such as “I” and “eye” or “made” and “maid”, as well as near homophones, such as “her” and “or”. Also, given that neither the number of words nor the boundary of each word in the input waveform is known, some of the sentences include strings of three words (e.g. “I’m aid”, “I’m eight”) that match the two first distinct sequences of phonemes put together (/ai/ and /meid/); these triplets are instances of homophony as well.

All sentences in the second column of Table 2 are ungrammatical (in the sense that they cannot be generated by the rule-driven structure of the English language), nonsensical, or simply strange. These sentences are examples of ‘system’ ambiguities, since the utterance would not be regarded as ambiguous by a person (Hutchins & Somers, 1992). The distinction between
what might be considered as ambiguous by a computer as opposed to what might be considered as ambiguous by a person helps understand the challenges of language modelling.

Although this sample utterance has been deliberately concocted to illustrate the pervasive nature of linguistic ambiguity, it does not take much imagination to think of “less extreme” utterances containing multiple homophones that are likely to induce recognition errors.

A speech recognition engine that lacks a language model would return all strings in the second column of Table 2, and more, because it would limit itself to matching all distinct sequences of phonemes that fit in the input signal to all words associated with those sequences of phonemes in the pronunciation dictionary. What the language model does, as its name implies, is to determine the correct language, i.e., the words and the order of those words, according to a certain model.

Unlike the acoustic modelling task, which is virtually always statistical in nature, language modelling can be either statistical or rule-based, depending on the characteristics of the application domain for which the system is designed. This dichotomy—opting for a statistical or a linguistic approach to solve ambiguity problems—is at the very heart of all natural language processing.

Statistical models are based on the probability of an individual word being followed by another one in a computer-readable collection of example texts. Thus, they can also be referred to as data-driven or corpus-based. The central idea is very straightforward: “given N choices for some ambiguous input, [a statistical model will] choose the most probable one” (Jurafsky & Martin, 2009, p. 6).

This approach is formalized by counting the occurrence of contiguous sequences of words in a corpus. Two-word sequences, like “I made” or “made her”, are called bigrams; three-word sequences, like “I made her” or “made her duck”, are called trigrams; and so on and so forth.

Let us take as an example the ambiguous utterance presented in Table 2. In any corpus of the English language, the word “I” would be more likely to be followed by the word “made” than by the word “maid”, even if their pronunciation is the same. This type of information, which is purely statistical in nature, is then used to make generalizations about language and resolve ambiguity at different levels.
A quick contrast experiment run in Google Ngram Viewer⁢ proves this. Figure 9 shows that the occurrence of the bigram “I made” is much more frequent than the occurrence of the bigram “I maid” in a large corpus of English books published across two centuries (the diachronic aspect is not actually relevant to the language modelling task; this exercise is merely intended to illustrate the difference in frequencies between the two bigrams).

![Figure 9. Google Ngram Viewer.](image)

The most widely-used technique deployed in distinguishing acceptable from unacceptable strings of words is based on highly complex probabilistic models called Hidden Markov Models (HMM), which assign a label or class to each unit in a sequence and then compute a probability distribution over possible labels in order to choose the best label sequence (Jurafsky & Martin, 2009).

HMMs can be applied to model sequences of words, but also to model units at the sub-word level, such as morphemes and phonemes. In fact, the acoustic modelling phase described in the previous section, by which phonemes can be recognized in the visual representation of a sound wave, is based almost entirely on this statistical model.

Using HMMs to determine the probability distribution of various linguistic units dates from the 1970s and was a major breakthrough in speech technology at the time. Nonetheless, a newer statistical technique developed in recent years, called Deep Neural Networks, improved recognition rates by over 30% (Seide et al., 2011). Word error rate in speech recognition went from one word in 4 or 5 words, to one word in 7 or 8. This technique has had a greater impact

---

³ The [Google Ngram Viewer](https://books.google.com/ngrams) is a free online application that shows the relative frequencies of n-grams over time.
on the performance of acoustic models, but it has also been proven to enhance the performance of language models (Arısoy et al., 2012).

At present, statistical language models are the most popular choice for commercial ASR systems and are better suited for large-vocabulary applications. This is mainly due to the vast increases in processing power and the availability of electronic corpora. However, they present clear disadvantages over rule-based approaches.

The most salient drawback is that they need a substantial amount of data from which to extract the minimum statistics that are required to model the words. Commercial recognition engines based on statistical language models typically need up to tens of thousands of utterances (Chen & Jokinen 2010, Chapter 12).

Conversely, so-called rule-based language models need much fewer example data or, in some cases, none at all. This makes them the most attractive choice for applications that target newer domains and involve under-resourced languages. As their name implies, these models focus on developing a set of rules that constrain the search space in a linguistically motivated way.

The most common way of doing this is by writing a Context Free Grammar (CFG) (Jurafsky & Martin, 2009), a formalism used for modelling natural languages based on the notion of constituent structure. Constituents are words or groups of words that behave as individual units within a hierarchical structure. Other formalisms that are based on this idea are the head-driven phrase structure grammar (HPSG) and the lexical functional grammar (LFG).

CFG grammars, which can also be referred to as Phrase-Structure Grammars, are made of a first set of rules intended to describe the syntactical structure of a language (a component that might be considered as a grammar in itself), and a second set of rules that is made up of a list of words and symbols.

For example, to account for the behavior of noun phrases in the English language, a CFG grammar would include a first set of rules stating that a noun phrase can consist of a noun, a determiner followed by a noun, an adjective followed by a noun, a noun followed by another noun, a pronoun, etc.; and a second set of rules —the lexicon— that links the actual words of the language to their grammatical categories (expressed as symbols in the grammar rules).
Table 3 below provides a few examples of what the rules that make up the grammar and the lexicon components of a CFG would look like:

<table>
<thead>
<tr>
<th>Context Free Rules</th>
<th>Grammar</th>
<th>Lexicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP → Det N</td>
<td>Det → a</td>
<td></td>
</tr>
<tr>
<td>NP → Adj N</td>
<td>Det → the</td>
<td></td>
</tr>
<tr>
<td>NP → N</td>
<td>N → car</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adj → fast</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adj → blue</td>
<td></td>
</tr>
</tbody>
</table>

*Table 3. Context free rules.*

This small toy grammar could generate a total of five noun phrases: “a car”, “the car”, “fast car”, “blue car”, and “car”, which in full would constitute the context-free language, i.e., “the set of strings that can be derived from a context-free grammar” (Jurafsky & Martin, 2009). Now, one would naturally want to increase as much as possible the number of strings that the CFG can generate in order to address real-life communicative situations, and that is when things get complicated.

In the example shown in Table 3, the relation between the number of context-free rules and the number of utterances that can be generated by those rules is not efficient at all. When writing a CGF language model, developers seek to reduce the number of rules and increase the size of the context-free language as much as possible.

The problems related to streamlining the rules of a context-free-grammar constitute the biggest challenge of rule-based language models: development is an arduous and time-consuming task that requires expert knowledge about linguistics and computer science, even for applications that target fairly limited domains.

One way of writing CFG grammars in a more compact way is by attaching pairs of features and values to the main elements that make up the rules, and then combining those elements in an efficient way so as to blend several rules into one. The idea is based on verifying the compatibility of values to ensure, for example, that an adjective and a noun have the same grammatical gender. These formalisms are known as feature or unification grammars (Rayner
et al., 2006). Another way of making a formalism more compact is by writing regular expression grammars. This is the approach used in the architecture of AidSLT and will be explained in detail in Section 3.4.

1.3 Conclusion

With this in mind, we conclude the present section by characterizing the language model deployed in AidSLT as a rule-based language model (since it will only try to recognize strings that are covered by the grammar) and we anticipate one distinguishing feature of the tool that contrasts with the main disadvantage mentioned above, namely: that the development efforts required for implementation are relatively moderate.
Chapter Two

II. Machine Translation

The overall goal of this chapter is to lay the groundwork for discussing the most salient architectural features associated to the translation strategy deployed in AidSLT, which will be briefly introduced at the end of this chapter and fully dealt with in Chapter III.

This chapter is organized as follows. Section 2.1 deals with the main problems involved in the automation of the translation process. Section 2.2 describes the main approaches to Machine Translation, which are presented in two broad groups that correspond to the type of information they rely on to solve translation problems: linguistic and nonlinguistic. Section 2.3 presents a few reflections about the integration of MT technology in a speech translation system. Finally, Section 2.4 concludes by presenting a summary account of the whole speech translation process and listing some characteristics of the translation strategy deployed in AidSLT.

2.1 Translation problems

Machine translation systems have to deal with all problems of ambiguity in the source language (except for homophones, in the case of MT systems for written input), as well as all problems that result from lexical and structural differences between the source and target languages.

Lexical differences between languages can be thought of as transfer or translational ambiguities. These arise “when a source language word can potentially be translated by a number of different target language words or expressions, not because the source language word is ambiguous, but because it is ambiguous from the perspective of another language” (Hutchins & Somers, 1992).

Most differences of this sort stem from conceptual mismatches between natural languages. For instance, the word fish can be translated into Spanish by the words pez or pescado depending on whether the animal is alive or dead. Other examples include wall, which can be rendered in German as Wand or Mauer depending on whether it refers to the inside or the outside of a wall;
and leg, which in French will translate as jambe, patte or pied depending on whether it belongs to a person, an animal or a table.

Differences in the lexical systems of languages can also stem from grammatical or stylistic features. An example of the first case is that the translation into Portuguese of the verb to know will vary depending on whether the direct object of the verb is a noun phrase or a subordinate clause. Thus, the sentence “I know the city” will be rendered into Portuguese as “Conheço essa cidade”, whereas “I know where to go”, will translate into “Sei para onde ir”. Stylistic mismatches are more common between languages that do not share a cultural background. Japanese speakers use different words for kinship relations depending on whether the kin they are referring to is their own or the hearer’s. Thus, kanai means ‘my wife’, and okusan ‘your wife’ (Hutchins & Somers, 1992).

Structural differences, for their part, occur when the semantic content of a source language sentence is conveyed in a target language sentence that is morphologically or syntactically different. These are much more varied and complex than lexical differences and hence more difficult to categorize.

One category established to describe structural differences between English and French (Vandooren in Bouillon et al., 1993) encompasses all cases where the same complement (or ‘argument’, an expression that completes the meaning of a predicate) is realized by different syntactic groups in the source and target languages. For example, the English sentence “I’m waiting for the postman to pass”, will translate in French as “J’attends le passage du facteur”. In English, the arguments of the verb to wait are a prepositional phrase followed by an infinitive verb (V + PP + Infinitive), whereas in French, the verb attendre takes a noun phrase and a prepositional phrase (V + NP + PP).

Another example is when the meaning of a word in the source language is lexicalized in the target language by a word of a different grammatical category. For example, the semantic load of the adjective ‘homeless’ in the phrase “homeless people” is conveyed in Spanish by a noun (“personas sin hogar”).

2.2 Main approaches to MT

Building an automatic engine that produces high-quality translations of random domain-independent texts is still far from being accomplished. The research community at large, both
at universities and big language-processing companies, treats this as a long-term goal, and many consider it to be simply impossible.

Thus, current machine-translation systems usually settle for less ambitious objectives. In general, they put the focus on: 1) tasks for which a rough translation is adequate; 2) tasks where a human can post-edit MT output; and 3) tasks limited to small sublanguage domains in which fully automatic high-quality translation is deemed achievable (Jurafsky & Martin, 2009).

Among the very first basic design questions that must be addressed in building an MT system are the number of languages that the system will cover and the translation direction for each language pair. Bilingual systems translate between a single pair of languages, whereas multilingual systems translate among more than two. Similarly, unidirectional systems are designed to translate from one language into another, whereas bidirectional systems can translate from and to both members of a language pair (Hutchins & Somers, 1992).

Machine translation systems try to solve the problems described in the previous section by turning to two types of information: linguistic (phonologic, morphologic, syntactic and semantic) and non-linguistic (statistics drawn from human-translated parallel texts). This is pretty much the same global choice developers are faced with when building a language model for a speech recognizer, as described in Section 1.2.3.

**Linguistic or ruled-based MT systems** may be classified as direct or indirect depending on the characteristics of their overall processing organization. As we will soon see, the main difference between these two broad approaches lies on how thoroughly they analyze the source text: direct systems conduct a minimal, local analysis, whereas indirect systems analyze the source language with greater depth.

The **direct approach** is the earliest historically; MT systems that are based on this method are referred to as “first generation”, they are typically built to cover one pair of languages and translate in one direction only. Their architecture is quite rudimentary. Very roughly, they run a lexical analysis of the input text (mainly to reduce inflected variants to canonical forms) and disambiguate the words syntactically; then they look all the words up in a large bilingual dictionary; and, finally, they apply some reordering rules to the target language words. They do not attempt to conduct a syntactic or semantic analysis of the source language sentences. Despite its limited capabilities, this approach is still valid today because it can offer various
advantages depending on the users' needs. The Météo system\(^4\), for instance, relies on a semi-direct approach.

The *indirect approach* appeared as a consequence of the poor performance of first generation systems, and in particular as a challenge to the premise that a translation of acceptable quality can be achieved simply by “molding” the source text into a target language output. These systems, which came to be known as the second generation in MT, are based on the idea that the automation of the translation process requires more sophisticated knowledge about the source and target languages. (Hutchins & Somers, 1992)

Indirect systems conduct a deeper analysis of the source phrase and come up with some kind of linguistic representation that is needed for translation. For the analysis of the source text and the generation of the target text, they rely on separate linguistically-motivated formalisms that describe all acceptable phrases in each respective language. These formalisms are called grammars, and were already introduced in Section 1.2.3 when discussing rule-based language models. Two variants of this approach, the transfer and the interlingua methods, will be briefly discussed next.

The *transfer* method can be divided into three phases: analysis, transfer and generation. First, the system conducts an analysis of the source language phrase and creates a language-dependent syntactic representation (typically a parse tree). Second, it applies a set of transfer rules to convert the source language representation into a target-language one. These transfer rules, which are specific to the source and target languages, are intended to resolve the lexical and structural differences between languages described earlier in the section. Finally, the system uses the target language grammar to generate a phrase out of the target-language representation. (Hutchins & Somers, 1992)

The *interlingua* method produces a language-independent semantic representation of the source phrase (called interlingua), and then generates a target-language phrase out of that representation without additional steps in between. Thus, it proceeds in two phases instead of three: from the source language to the interlingua, and from the interlingua to the target language. The analysis conducted by the system must be exhaustive enough so that the language-independent representation contains all the information that is needed to generate the

\(^4\) A well-known machine translation system designed in the 70’s by the University of Montreal to translate weather forecasts.
target phrase (this is the biggest challenge). This strategy is clearly more attractive for multilingual systems, since it does away with the need to build transfer modules for each new language and each new translation direction.

The pyramid in Figure 10 illustrates the three types of MT systems described above according to the depth of the analysis they require, going upwards from the base to the top. Direct systems lie at the bottom, transfer systems in the middle, and interlingua systems at the top.

![Figure 10. Linguistic MT systems (Hutchins & Somers, 1992).](image)

Finally, the other broad general approach to machine translation consists in extracting statistical information from human-translated parallel texts and using it to ‘teach’ a computer how to translate. These MT systems are known as statistical, data-driven or corpus-based. This approach has been the most popular choice for quite some time now due to vast increases in computers’ processing power and the availability of multilingual electronic corpora.

The main mathematical models and statistical techniques deployed by this type of MT system are pretty much the same as the ones described in earlier sections. They rely on the distributional properties of words and phrases to establish the most likely translation of the input text.

They do this, very roughly, in two steps: they first extract information from a parallel corpus to find the most adequate translation of a given sequence of words in the source language, and then they resort to a language model like the one described in Section 1.2.3, fed by monolingual target-language data, to “analyze the linguistic correctness of the output translation text, choose
the best possible linguistic combination and optimize fluency in the target” (Porro Rodriguez, 2012).

2.3 SLT Systems

This master’s dissertation began by saying that SLT is especially difficult because it involves dealing with all fundamental problems associated to each of its individual component technologies, but also because integrating those technologies presents additional problems that are very hard to solve (see Section 1.1). Particularly, the biggest challenge lies on combining speech recognition and machine translation in an efficient way.

The majority of current SLT systems do this by submitting a string of text produced by a speech recognizer as input to some kind of machine translation engine. In the early nineties, when the field of SLT was taking its very first steps, the most popular choice was building pure linguistically-motivated systems. However, the limitations of this approach, along with the increasing availability of large electronic corpora and processing power, led to the development of hybrid architectures that combine both data-driven and rule-based aspects. At present, most SLT systems are based on hybrid architectures (Chen & Jokinen 2010, Chapter 12).

Other systems simply do away with the translation component of the system and limit themselves to matching spoken words against a set of fixed phrases. This approach to SLT, which is often referred to as “fixed-phrase translation”, favors accuracy over robustness and is typically deployed in safety-critical applications.

One commercial system that deploys this method is the Phraselator (Sarich, 2004). This system is defined by its developers as a “speech actuated phrase matcher” (p. 3), rather than a proper speech translator. It picks a phrase or a sequence of phrases based on the input signal and plays a recording of the target-language output (it also does away with speech synthesis). The Phraselator translates from English into other languages and is intended for use in military settings.

Another example is MediBabble, a fixed-phrase portable translation app that has achieved considerable popularity with medical practitioners. It contains a large multilingual database of medical questions and instructions. This tool is even less sophisticated than the Phraselator,

---

5 medibabble.com (accessed in August 2015)
since it also does away with speech input. Users just need to press on the question they want to ask and the app plays a recording of that question in the target language. It is currently available in five languages and can be deployed without internet connectivity.

MediBabble cannot be said to be a proper SLT system, since it does not handle speech input and it does not synthesize output. Nonetheless, it actually plays the same role as proper SLT applications in that it allows two people who do not speak the same language to communicate orally.

One interesting aspect of these systems is that they were both developed by domain experts. The Phraselator was commissioned by the Defense Department of the United States and developed by a former soldier (Chen & Jokinen 2010, Chapter 12). Similarly, the content in MediBabble was developed by group of doctors. In the case of MediBabble, the fact that content was developed by domain experts is used to advertise the product. The first feature displayed on the ‘Features’ tab on the web site of the application says that content was “written and reviewed by a panel of physicians supervised by Dr. Lawrence Tierney, Professor of Medicine at UCSF School of Medicine and author of the seminal text on evidence-based history-taking”6.

Another distinguishing feature is that content is organized in a modular structure that exploits generic aspects of the domain. In the case of MediBabble, for example, questions and instructions are grouped by care provider and specific subdomains such as “History of Present Illness”, “Past Medical History” or “Medications and Allergies”, among other relevant categories. This is done to facilitate navigation through the database.

Figure 11. The Phraselator and MediBabble.

---

6 medibabble.com (accessed in August 2015).
2.4 Conclusion

The process of translating spoken language can be summed up in the following way: the acoustic model and the pronunciation dictionary suggest different possible sequences of words based on the input speech signal (this is done following a statistical model); then the language model weighs them up and picks one acceptable sequence of words (this can be done according to a statistical model or a set of linguistic rules), and finally a translation engine (which can also be statistical or rule-based) produces a target language version. In the case of speech-to-speech systems, the target language output is processed by a voice synthesizer.

AidSLT is presented as a multilingual unidirectional speech translation system for limited domains. Its architecture allows for more expressivity than fixed-phrase systems but less than general syntax-based transfer or interlingua architectures. The strategy deployed by the tool to translate spoken input will be explained more in detail in the following chapter.
Chapter Three

III. AidSLT

The objective of this chapter is to take a closer look at the tool, which previous chapters have only sketched. This chapter describes the first prototype of AidSLT and discusses several aspects related to its context of use and its development. The reader will be presented with an overview of the software development process, including how it was conceived, the criteria for defining content, the needs that were identified according to the context of use, and how those needs were later rendered into specific functionalities in the formalism and the user interface. Part of the material presented in this chapter, especially in the sections dealing with the corpus, the architecture and the user interface, has been already published in the proceedings of the 36th session of the ASLIB Conference (Armando et al., 2014).

Section 3.1 examines the rationale for developing the tool. Section 3.2 defines the intended context of use and Section 3.3 explains the criteria that were used to build the corpus, i.e., to define the content for the first prototype. Section 3.4 describes the architecture of the system and Section 3.5 presents the tool from the user’s point of view. Section 3.6 concludes.

3.1 Motivation

It is often the case that international civil servants deployed in field operations do not share a common language with the beneficiaries of humanitarian aid. In environmental emergencies, for example, humanitarian experts are sent to the field within hours after the onset of a disaster to assess the needs of the affected population, where they usually encounter difficulties in carrying out their activities because they do not speak the local language.

The United Nations Office for the Coordination of Humanitarian Affairs (OCHA) publishes an annual report that intends to provide a comprehensive picture of the global humanitarian landscape and to highlight the major trends, challenges and opportunities concerning humanitarian crisis and assistance. In their most recent report, OCHA lists language barriers as one of the major operational challenges in 2013.
“Communicating with affected communities is a challenge for international aid workers. In 2013, there was an average of 59 living languages (or dialects) spoken in countries with an appeal. In six countries (Afghanistan, Mauritania, occupied Palestinian territory, Somalia, Syria and Yemen), the language of official UN correspondence was not the same as the country’s official language(s). The language barrier also extends to oral communications, creating a need for international humanitarians to work closely with local partners to ensure that communication channels with affected communities remain open.” (OCHA, 2014, p. 32)

Moreover, the regions of the world that are most affected by poverty, conflicts and natural disasters also happen to be the most linguistically diverse, and their populations often speak under-resourced languages.

The Greenberg’s Diversity Index measures the probability that two people selected from the population at random are native speakers of the same language; it ranges from 0 (everyone has the same mother tongue) to 1 (no two people have the same mother tongue). The vast majority of the countries that rank the highest in that list are typical beneficiaries of humanitarian aid. In the next section, we will see how this applies to the specific context of use targeted by AidSLT.

3.2 Context of use

AidSLT is intended to be used by humanitarian workers to administer multilingual household surveys about malaria prevention measures. The present section provides a number of details about the application domain, and Section 3.3 below describes the specific criteria that were used to build the corpus.

Malaria poses a major public health problem across the globe. It is estimated that 3.3 billion people, or 40 percent of the world’s population, live in areas of malaria risk. Approximately 198 million malaria episodes and 584,000 malaria deaths occurred in 2013. Young children and pregnant women represent those at greatest risk of being infected (World Health Organization, 2014).

While malaria is endemic within most tropical and subtropical regions of the world, over 90 percent of all malaria deaths currently occur in sub-Saharan Africa, which incidentally is one of the most linguistically diverse regions of the world. The 18 countries that registered the
highest number of deaths caused by malaria in 2013\textsuperscript{7} score an average of 0.833 in the Greenberg’s Diversity Index. For the European reader to have a closer reference point, Switzerland, a very small country with four official languages, scores 0.547 in that same list.

One of the most effective strategies to fight malaria is vector control, i.e., taking preventive measures to eradicate the mosquito that transmits the disease. The two core, broadly applicable malaria vector control measures are distributing insecticide treated nets (ITNs) and spraying the interior walls of houses against mosquitoes. The latter practice is known as indoor residual spraying (IRS).

Both of these measures have a proven track record of successfully reducing or interrupting disease transmission, particularly in areas that are highly prone to malaria. ITNs alone are estimated to have reduced malaria mortality rates by 55 percent in children under five years of age in sub-Saharan Africa (World Health Organization, 2014).

For this reason, the governments of malaria-endemic countries, with the help of NGOs and international humanitarian organizations, adopt specific policies to protect their populations from the disease, such as distributing free mosquito nets and conducting IRS campaigns, and then they monitor the impact of those policies by means of household surveys. Household surveys are a principal measurement tool to collect data for measuring indicators for malaria control (World Health Organization et al., 2013).

Basically, humanitarian organizations and governments send teams of interviewers to people’s houses to collect data about concrete preventive measures, such as how many mosquito nets the family owns or whether their house has already been sprayed with insecticide. \textit{This is the intended context of use for AidSLT}: an interviewer, working in a team with other interviewers —as is usually the case for this type of surveys— administering a number of very specific questions that come with a set of predefined answers.

Interviewers have traditionally used paper-based questionnaires, but in recent years organizations have started to use handheld devices, such as mobile phones or tablets. Mobile data collection is becoming more and more popular in humanitarian settings, since it has proven to be faster and more cost-effective. Two examples are the Rapid Mobile Phone-based (RAMP)\textsuperscript{7}\textsuperscript{7}.

---

\textsuperscript{7} Uganda, Ghana, Niger, Chad, Mozambique, Burkina Faso, Ethiopia, Côte d’Ivoire, Mali, Guinea, Cameroon, Kenia, Zambia, Malawi, Benin, Senegal, Indonesia and Central African Republic (WHO, 2014).
surveys\textsuperscript{8} conducted by the International Federation of Red Cross and Red Crescent Societies (IFRC), and the Standardized Expanded Nutrition Surveys\textsuperscript{9} (SENS) conducted by the refugee agency of the United Nations (UNHCR).

This type of oral exchange seems reasonably attractive for a rule-based SLT device: it has a limited vocabulary; it presents a moderate level of syntactic complexity (neither fixed phrases nor unrestricted speech); there is not enough data available to train a statistical engine; it is a type of task-oriented collaborative dialogue where there is a marked asymmetrical relationship between participants (interviewer is in control of the exchange); and precision is more important than robustness. In this sense, it is analogous to other application domains targeted by similar tools, such as MedSLT, the Phraselator and MediBabble (see Section 2.3).

### 3.3 Corpus

The corpus is made up of 63 questions intended to measure 9 internationally recognized vector control indicators developed by a cluster of relevant international organizations (World Health Organization et al., 2013). Table 4 below lists all nine indicators.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Indicator Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevention</strong></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Proportion of households with at least one ITN</td>
</tr>
<tr>
<td>2.</td>
<td>Proportion of households with at least one ITN for every two people (NEW)</td>
</tr>
<tr>
<td>3.</td>
<td>Proportion of population with access to an ITN within their household (NEW)</td>
</tr>
<tr>
<td>4.</td>
<td>Proportion of population that slept under an ITN the previous night</td>
</tr>
<tr>
<td>5.</td>
<td>Proportion of children under five years old who slept under an ITN the previous night</td>
</tr>
<tr>
<td>6.</td>
<td>Proportion of pregnant women who slept under an ITN the previous night</td>
</tr>
<tr>
<td>7.</td>
<td>Proportion of existing ITNs used the previous night (NEW)</td>
</tr>
<tr>
<td>8.</td>
<td><em>Households covered by vector control</em>: Proportion of households with at least one ITN and/or sprayed by IRS in the last 12 months</td>
</tr>
<tr>
<td>9.</td>
<td><em>Universal coverage of vector control</em>: Proportion of households with at least one ITN for every two people and/or sprayed by IRS within the last 12 months (NEW)</td>
</tr>
</tbody>
</table>

*Table 4. Vector Control Indicators (World Health Organization et al., 2013, p. 7)*

\textsuperscript{8} \url{ifrc.com/ramp} (accessed in August 2015)

\textsuperscript{9} \url{sens.unhcr.org} (accessed in August 2015)
Some of the questions are about malaria prevention measures — namely, the use of mosquito nets and indoor residual spraying — and bear a more immediate relevance to the indicators in Table 4, while others are about household characteristics, such as water and sanitation facilities, and are needed to establish a connection between the socioeconomic status of the household and public health issues.

All questions were taken from two model household surveys recommended by the major humanitarian actors that developed the indicators, which include the United Nations Children’s Fund (UNICEF), the World Health Organization and the Roll Back Malaria Partnership. The questionnaires are the Multiple Indicator Cluster Survey (MICS)\textsuperscript{10} and the Malaria Indicator Survey (MIS)\textsuperscript{11}.

MICS surveys were developed by UNICEF to support countries in monitoring the situation of children and women. These surveys are designed to produce data that are comparable over time and across countries and are harmonized with other major household survey programs, such as the MIS survey described below. The MICS survey has a modular structure that allows designers to include only relevant modules in their questionnaires. The package includes a malaria module that can be used to collect all necessary data for measuring the indicators in Table 4.

The MIS survey was developed by the Roll Back Malaria Partnership, a global framework for coordinated action against malaria\textsuperscript{12}. The scope of this survey is much more limited, since it has been designed specifically for collecting data about malaria. Like the MICS survey, the MIS survey is structured by modules and allows for collecting all necessary data for measuring the indicators in Table 4.

The two figures below show screenshots from one of the many questions that are repeated in both paper-based questionnaires, so the reader can see what they look like. The question reads “What is the main source of drinking water for members of your household?”.

Figure 12 shows a screenshot from the MICS questionnaire, and Figure 13 shows how the same question looks like in the MIS questionnaire. In both cases, the question is the first of the water and sanitation module (the code is “WS1” in MICS, and 101 in MIS) and comes with a

\textsuperscript{10} mics.unicef.org (accessed in August 2015)
\textsuperscript{11} malariasurveys.org (accessed in August 2015)
\textsuperscript{12} rollbackmalaria.org (accessed in August 2015)
predefined set of answers. The third column in Figure 12 and the fourth in Figure 13 specify what the following question should be according to the answer obtained.

**Figure 12. The MICS survey.**

<table>
<thead>
<tr>
<th>NO.</th>
<th>QUESTIONS AND FILTERS</th>
<th>CODING CATEGORIES</th>
<th>SKIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>What is the main source of drinking water for members of your household?</td>
<td>PIPED WATER</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIPED INTO DWELLING</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIPED TO YARD/LOT</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TUBE WELL OR BOREHOLE</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DUG WELL</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROTECTED WELL</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UNPROTECTED WELL</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WATER FROM SPRING</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROTECTED SPRING</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UNPROTECTED SPRING</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAINWATER</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TANKER TRUCK</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CART WITH SMALL TANK</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SURFACE WATER (RIVER/DAM)</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BOTTLED WATER</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OTHER</td>
<td>96</td>
</tr>
</tbody>
</table>

**Figure 13. The MIS survey.**
3.4 Architecture of the system

AidSLT was built using Regulus Lite\textsuperscript{13}, an open source platform for building spoken language translation applications that supports rapid development and web deployment of several types of phrasal speech translation systems using a minimal formalism, which will be described in detail in the present section.

The description of the application is compiled into an accurate limited-vocabulary speech recognizer and a set of tables. Applications can be hosted on mobile platforms — smartphones, tablets or laptops — linked over a 3G connection to a remote server with recognition performed on the server (Fuchs et al., 2012). It is deployed over the web using methods developed at Geneva University and elsewhere in the context of various speech translation and CALL projects (Rayner et al., 2006; Fuchs et al., 2012).

The questionnaire designer specifies the questionnaire using a single file written in a simple formalism which supports three types of unit: Groups, Questions, and Answers.

A Group specifies a top-level item on the questionnaire, a list of permitted Fillers, and a pointer to the next Group. A Question specifies one possible way to attempt to assign a Filler to a Group; it defines the Group which the Question belongs to, a list of surface realizations of the question in the questionnaire administrator’s language, a translation into each target language, and a list of permitted Answers, each one optionally associated with a Filler. An Answer defines an associated translation of a given filler in each target language.

Figure 14 below illustrates a Group called WaterForDrinking, which was included in the questionnaire administered by evaluators in the simulation exercise. Figures 12 and 13 in Section 3.3 show how this question looks in the original paper-based questionnaires that were used to build the corpus.

\begin{verbatim}
Group
Questionnaire EVALUATION
Name WaterForDrinking
PrintName What is the main source of drinking water for members of your household?
Code 7
\end{verbatim}

\textsuperscript{13} \url{www.issco.unige.ch/en/research/projects/LiteDocSphinxTrans/build/html/development.html} (accessed in August 2015)
HelpFile 7.xml
Fillers piped_into_dwelling piped_to_yard/plot piped_to_neighbour
public_tap/standpipe tube_well_or_borehole protected_well
unprotected_well protected_spring unprotected_spring
rainwater_collection tanker_truck cart_with_small_tank surface_water
bottled_water other
Next SourceLocation
EndGroup

Figure 14. Group 7. WaterForDrinking.

The other fields have the following meanings. PrintName is the title displayed at the top of the screen when the Group is reached. This is the “visible face” of the Group, so to speak, and it always matches a question or a command in a paper-based questionnaire.

Code is an identifier used to store the results when the questionnaire is filled. The idea of adding codes to Groups and Answers was included to facilitate the processes of sending and processing the collected data. In an emergency situation, for example, it might be useful to send data in the form of phone text messages.

HelpFile associates a Group to a set of question-specific guidelines that are meant to help the interviewer understand the exact purpose of what they are asking. These guidelines are stored in the form of an XML file and displayed in the INSTRUCTIONS tab of the tool (see Section 3.5). The content has been adapted from specific manuals published by the humanitarian organizations that developed the paper-based questionnaires. The box below shows an extract of the help file associated to the Group described above.

<table>
<thead>
<tr>
<th>WHAT IS THE MAIN SOURCE OF DRINKING WATER FOR MEMBERS OF YOUR HOUSEHOLD?</th>
</tr>
</thead>
<tbody>
<tr>
<td>The purpose of this question is to assess the type of household water used for drinking. Definitions of the various sources of water are as follows:</td>
</tr>
<tr>
<td>PIPED INTO DWELLING, also called a house connection, is defined as water service connected by pipe with in-house plumbing to one or more taps, for example, in the kitchen and/or bathroom.</td>
</tr>
<tr>
<td>PIPED INTO COMPOUND, YARD OR PLOT, also called a yard connection, is defined as a piped water connection to a tap placed in the compound, yard or plot outside the house.</td>
</tr>
<tr>
<td>[...]</td>
</tr>
<tr>
<td>A TUBE-WELL OR BOREHOLE is a deep hole that has been driven, bored or drilled with the purpose of reaching groundwater supplies. Boreholes/tube-wells are constructed with casing, or pipes, which prevent the small-</td>
</tr>
</tbody>
</table>
diameter hole from caving in and provide protection from infiltration of run-off water. Water is delivered from a tube-well or borehole through a pump that may be powered by humans, animals, wind, electricity, diesel fuel or solar energy.

There are fifteen possible ways to fill the WaterForDrinking slot: piped_into_dwelling; piped_to_yard/plot; piped_to_neighbour; public_tap/standpipe; tube_well_or_borehole; protected_well; unprotected_well; protected_spring; unprotected_spring; rainwater_collection; tanker_truck; cart_with_small_tank; surface_water; bottled_water; and other. These are specified in the Fillers field.

Once the slot has been filled, the questionnaire moves to the Group SourceLocation (“Where is that water source located?”). However, the Group definition under analysis is actually a simplified version of the original one, which had to be adapted to fit the simulation exercise designed for the evaluation. In the original corpus (described in Section 3.3), as in a real-life situation, the interviewer would obviously skip the question about the location of the source if the respondent chooses “piped_into_dwelling”.

The Next field is intended to reproduce the original sequence structure of the paper-based questionnaire, where the next question is always determined by the answer of the question at hand. In paper-based questionnaires, the sequence structure is typically specified with arrows and codes, as shown in the third column of the Multiple Indicator Cluster Survey (see Figure 12) and in the fourth column of the Malaria Indicator Survey (see Figure 13). To reflect the sequence of questions, the designer of the questionnaire can add additional Next lines and use conditionality operators, as illustrated below:

```
Next WaterTreatment if piped_into_dwelling; piped_to_yard/plot;
piped_to_neighbour

Next SourceLocation if public_tap/standpipe; tube_well_or_borehole;
protected_well; unprotected_well; protected_spring;
unprotected_spring
```

A Group will normally have several Questions associated with it; indeed, the value of the tool, compared to an application which mechanically asks a set of questions in a fixed order, is that
it allows human judgment to be used to select an appropriate question. In the current case, the interviewer might want to ask a clearer, more specific question if the interviewee is not sure about what to answer, such as “Does the water that you drink come from a tap in your house?”.

Figure 17 below shows an example of a Question definition.

---

**QuestionTemplate** yes_no_drinkingwater FRENCH SPANISH ENGLISH VALUE

**Group** WaterForDrinking

**Translation/French** L'eau que boivent les membres de votre ménage vient-elle FRENCH ?

**Translation/Spanish** ¿El agua que beben los miembros de su hogar viene SPANISH?

**Variant** do you ?{ principally | mainly | mostly | usually } ( get | collect ) ( your | the ) ?{ drinking ) water ?{ that you drink )

**Variant** does ( your | the ) ?{ drinking ) water ?{ that you drink ) ?{ principally | mainly | mostly | usually } come ENGLISH

**Answers** yes=VALUE no

End**QuestionTemplate**

---

**ApplyTemplate** yes_no_drinkingwater "d'un robinet dans la maison" "de un grifo en la casa" "from a tap ( in | inside ) ( the house | the kitchen | the bathroom )" "piped_into_dwelling"

---

**Figure 17. QuestionTemplate. WaterForDrinking.**

Here, the **Variant** lines define different ways for the (English-speaking) interviewer to ask the question, the **Translation** lines define how it will be asked in the different target languages covered by the system, and the **Answers** line lists possible responses. As shown, it is possible to use regular expressions in the **Variant** lines in order to list different forms more compactly.

In the definition displayed in Figure 17, the designer specified two **Variant** lines to cover for two alternative ways of asking the same thing (broadly, “Do you get your drinking water from a tap?”; and “Does your drinking water come from a tap?”), and used a number of regular expression operators to provide the user with some degree of lexical and syntactic reformulation. Thus, this single Question unit contains two broad alternative questions, which can be phrased in several dozens of ways, and that will be associated to one canonical form in each of the target languages.
Another resource to make the file more compact is to use templates and template applications. In the first line of Figure 17, the designer specified a name for the QuestionTemplate (yes_no_drinking_water) followed by variables for each of the languages covered by the system and a common value at the end. Then, in the associated template application, the designer assigned a variant for each of the languages and a filler for the common value. The correspondence between variants and variables is assigned by the order of the elements, as follows: “d'un robinet dans la maison” for FRENCH; "de un grifo en la casa" for SPANISH, "from a tap ( in | inside ) ( the house | the kitchen | the bathroom )" for ENGLISH; and the filler “piped_into_dwelling” for VALUE. As we can see, regular expressions are also permitted in the formulation of variants in template applications.

Naturally, the advantage of this resource is that the designer can associate additional template applications to the same template. Below are some examples from the evaluation file:

```
ApplyTemplate yes_no_drinkingwater "d'un robinet public" "de un grifo público" "from a ( public tap | standpipe )" "public_tap/standpipe"
ApplyTemplate yes_no_drinkingwater "d' un trou" "de un pozo" "from a ( tube well | borehole )" "tube_well_or_borehole"
ApplyTemplate yes_no_drinkingwater "d' une rivière" "de un río" "from a river" "surface_water"
```

*Figure 18. Template applications.*
This resource can be used only in some cases. Prepositional phrases are an example of constituents that can be treated in this way, but applying the same technique to other constituents immediately causes practical problems, revealing the lexical gaps and structural divergences between languages (see Section 2.1).

The degree of robustness of the tool, i.e., its capacity to effectively recognize variant input, is mostly determined by the pertinence of the content and the efficiency of the code (how well is written from a technical point of view), and its roof, so to speak, is the expressivity allowed for by regular expressions and the template mechanism described above.

This raises very interesting questions about the formalism. This application is designed so that domain experts can write the code themselves. As we have seen, it is fairly simple to write the formalism; it is based mainly on easy-to-use templates and regular expressions. The person who is most fit to define the content of the questionnaire is someone who knows the domain well, rather than someone who has knowledge of programming languages but is not familiar with the context of use.

Finally, the surface forms of the answers are defined by Answer units; a typical example is the following, which specifies how to realize the answer yes.

```
Answer
Content yes
PrintName Yes
Code 1
Translation/French Oui
Translation/Spanish Sí
EndAnswer
```

Figure 19. Answers.

3.5 User interface

When using the application, the first step is to choose the specific questionnaire the user wants to administer and the language of the interviewee. This is done in the SELECT QUESTIONNAIRE tab. If there were many questionnaires of different sorts, the user could choose to see them grouped by domain in the SELECT DOMAIN tab.
At the moment, the available choices are to administer malaria-relevant modules of the Multiple Indicator Cluster Survey (MICS) or the Malaria Indicator Survey (MIS) in French or Spanish. Adding other target languages more suitable for the intended context of application would only involve writing extra *Translation* lines in the relevant question definitions. This makes the tool rapidly portable to new target languages.

After choosing the questionnaire and the language, the user is directed to the operative tab of the application (USING THE APPLICATION), where they can start administering the questionnaire (Figure 20).

![A Prototype Form Filling Application](image)

Figure 20. USE THE APPLICATION tab.

The title at the top of the screen (What is the main source of drinking water for members of your household?) corresponds to a question in the original paper-based questionnaire and is defined in the *PrintName* field of the corresponding Group in the code. In a way, Figure 20 above completes the picture this chapter has tried to present so far: we have seen how a unit in a paper based questionnaire, composed of a question and a set of fixed answers (see Figures 12 and 13), was rendered into the code of an application designed for administering multilingual questionnaires (see Section 3.4), and now we present it from the user’s point of view.

The user asks a question by holding down the input button (top right) and speaking. When the user holds down the button, the background of the icon turns orange, signaling that the microphone is recording, and when the user releases the button, the background of the icon goes back to its default grey color and the system plays a recording of the user’s interaction. This
mechanism is known as “push-and-hold”; it is a microphone activation strategy by which the start and the end of speech is signaled by the press and the release of a button respectively (Tsourakis, 2013).

Just as in any conventional situation, the user can read the text at the top of the screen out loud or use their own judgment to rephrase the question or ask something else. If they are not sure about how to phrase their question or they have problems with the recognition, they have the option of pressing the Help button (bottom right) to get a list of example questions (bottom center).

When the interviewer has spoken the question, the recognition result appears on the screen so that the user can approve of it before eliciting the translation, which is displayed in the PROVIDE ANSWER tab together with all available answers (Figure 21).

As explained in the previous section, the application reproduces the sequence structure of the original questionnaire, so the next question is always determined by the answer of the question at hand. For instance, in the larger corpus, the application will automatically skip question HC12 in the MICS questionnaire (“How many hectares…?”) if the respondent gives a negative answer to HC11 (“Do you own any agricultural land?”). Similarly, the application will force the interviewer to probe for additional information whenever it is required by the paper based questionnaire.
Finally, as shown in the previous section, the INSTRUCTIONS tab contains question-specific guidelines for the interviewer.

3.6 Conclusion

This Chapter has presented a detailed picture of the rationale for the project, the intended context of use, the formalism and the user interface of the tool. It has made explicit all the features that were presented in previous chapters, including the degree of expressivity provided by the formalism and the translation strategy.

The most salient features that we can add to the ones presented in the previous chapters are that the tool favors rapid development by domain experts, and that it is easily portable to new target languages.
Chapter Four

IV. Evaluation

From March to July 2015, seven humanitarian workers with field experience and knowledge of household surveys tested the first prototype of AidSLT and answered a brief on-line questionnaire about its potential in humanitarian settings.

The goal of this chapter is to describe several aspects related to the design of the evaluation. Section 4.1 presents a number of methodological aspects associated to the evaluation of AidSLT. Section 4.2 details the three specific objectives that were set for the evaluation. Section 4.3 explains what the evaluation consisted of and describes the procedure that was followed to run the experiments, and Section 4.4 concludes.

4.1 Methodological aspects

The most relevant reference frameworks for designing an evaluation of a spoken language translation tool are to be found in the literature on software engineering, rather than in the literature on translation studies. The same applies to other language processing tools. The methodology proposed by Monika Höge to evaluate translators’ aids, for example, draws elements from various disciplines, but she acknowledges that software engineering is the one that has had the largest impact (Höge, 2002).

The methodology used to evaluate AidSLT is informed by the ISO 9126 standard, which concerns the quality of software products, and the final report written by the Evaluation Working Group of the Expert Advisory Group on Language Engineering Standards (EAGLES), which draws upon the ISO 9126 standard and defines a more specific methodology for the evaluation of language technology systems.

The ISO 9126 standard was originally developed in 1991 by the International Organization for Standardization (ISO) and perfected in several successive versions. It defines a reference quality model made up of six characteristics that a good software product should have:
functionality, reliability, usability, efficiency, maintainability and portability. These characteristics are meant to be broken down into measurable attributes and used as a basis for the design of evaluations. This standard has been used in the past to develop a quality model for the evaluation of similar speech translation tools (Tsourakis, 2013).

The EAGLES report was published in two parts; the first one appeared in 1996 and the second one in 1999. The Evaluation Working Group built on the ISO standard to establish a general framework for the evaluation of language engineering products, maintaining the central idea of building a hierarchical model to assess the quality of a software product. The report also contains valuable information about the various types of tests, testing techniques, methods and instruments. The six quality characteristics described in the ISO standard are also included in the EAGLES report, and are defined as follows.

**Functionality**
A set of attributes that bear on the existence of a set of functions and their specified properties. The functions are those that satisfy stated or implied needs. (ISO 9126: 1991, 4.1)

**Reliability**
A set of attributes that bear on the capability of software to maintain its level of performance under stated conditions for a stated period of time. (ISO 9126: 1991, 4.2)

**Usability**
A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users. (ISO 9126: 1991, 4.3)

**Efficiency**
A set of attributes that bear on the relationship between the level of performance of the software and the amount of resources used, under stated conditions. (ISO 9126: 1991, 4.4)

**Maintainability**
A set of attributes that bear on the effort needed to make specified modifications. (ISO 9126: 1991, 4.5)

**Portability**
A set of attributes that bear on the ability of software to be transferred from one environment to another. (ISO 9126: 1991, 4.6)
This set of characteristics can be used as an instrument to narrow down the scope of an evaluation by specifying the relative importance of each characteristic in a particular system. The evaluation of AidSLT focuses on the first of the six characteristics listed above since it is centered on the functional properties of the tool, that is, on determining whether the tool does what it is supposed to do.

The most suitable technique to test the functional properties of a tool—the one used in the evaluation of AidSLT—is called **black box testing**. This is a familiar concept in software engineering and it is typically defined in opposition to **glass-box testing**; the former looks at external elements, while the latter is more suitable to test the internal structure of a system. Black box testing is centered on user requirements; it examines “inputs causing anomalous behavior by the system” and “outputs which reveal the presence of defects” (Höge, 2002, p. 129). The evaluations designed according to this approach focus on overall performance and are typically conducted by end users.

The second part of the EAGLES report presents a seven-step “recipe” for the evaluation of language engineering products. The first two steps of the recipe include a number of standard questions intended to determine basic aspects of the system and the evaluation, such as why the evaluation is being done, what exactly is being evaluated, who will use the system, etc. It then includes three successive steps that reflect the central idea of the overall evaluation framework, i.e., defining a number of top level characteristics and breaking them down into sub-characteristics that eventually bottom out into metrics which can be applied directly to the software in question (the quality model). Finally, the recipe includes two last steps that relate to the design of the evaluation, which concerns the development of test materials and the definition of appropriate measurements, and its execution, which concerns taking the measurements and summarizing the results in a report.

In order to provide the reader with an overall picture of the test arrangements presented in successive sections, the evaluation of AidSLT is described against a simplified version of this recipe, which includes some information about the initial key definitions, the quality model and the design and execution of tests. The following sections of this chapter contain more detailed information about the various aspects of the evaluation.

Let us first consider some of the initial definitions. The main purpose of evaluating the first prototype of AidSLT was to determine the viability of the project, i.e., to determine if the tool
is likely to be deployed with some degree of success in its intended context of use. Hence, the evaluation can be categorized as an adequacy evaluation, which is defined as “the activity of assessing the adequacy of a system with respect to some intended use of that system” (EAGLES, 1996).

More specifically, what was evaluated is the ability of the tool to recognize speech input by English-speaking users and to translate it into French in the context of the evaluation task. These two abilities, recognizing and translating speech, can be considered as separate functionalities of the tool, as well as separate components of the system. Also, put together, they constitute the most important single attribute determining the overall performance of the tool (translating speech). Therefore, the object of the evaluation of AidSLT can be said to be a set of functionalities, a set of components and the system itself.

The tool is intended to be used by humanitarian workers to administer multilingual household surveys. As we shall see, the evaluation is in part intended to sketch the profile of humanitarian workers as potential users of SLT systems, especially rule-based systems like AidSLT, and to determine whether the intended context of use was a good choice. In fact, we recruited participants with field experience because we were interested in the opinion of humanitarian workers who have real contact with beneficiaries so they could confirm that language barriers are indeed a problem for them, and because we could ask them whether they consider household surveys to be an adequate context of use for a speech translation tool.

Regarding the quality model, as we have seen, the primary top level characteristic chosen to design the tests is the functionality of the tool. As explained above, once the top-level characteristics are defined, these have to be divided into sub-characteristics and attributes that can eventually bottom out into metrics which can be applied directly to the software in question. According to the ISO standard, functionality can be divided into five sub-characteristics: suitability, accuracy, interoperability, compliance and security. These are defined as follows.

**Suitability**

*Attribute of software that bears on the presence and appropriateness of a set of functions for specified tasks.* (ISO 9126: 1991, A.2.1.1)
Accuracy

Attributes of software that bear on the provision of right or agreed results or effects. (ISO 9126: 1991, A.2.1.2)

Interoperability

Attributes of software that bear on its ability to interact with specified systems. (ISO 9126: 1991, A.2.1.3)

Compliance

Attributes of software that make the software adhere to application related standards or conventions or regulations in laws and similar prescriptions. (ISO 9126: 1991, A.2.1.4)

Security

Attributes of software that bear on its ability to prevent unauthorized access, whether accidental or deliberate, to programs and data. (ISO 9126: 1991, A.2.1.5)

The sub-characteristics that were prioritized in designing the evaluation of AidSLT were suitability and accuracy. The definition of suitability in the ISO standard is quite broad: it is defined as an attribute of a software product that “bears on the presence and appropriateness of a set of functions for specified tasks” (ISO 9126: 1991, A.2.1.1). In the case of AidSLT, as we shall see in the following section, this attribute was associated with the ability of the tool to recognize variant input (robustness) and its ability to translate speech input in an adequate manner in the context of the evaluation task (task performance). Both aspects were measured by analyzing registered interactions along with their corresponding translations and annotating the data using binary metrics.

Accuracy was associated with the ability of the tool to recognize spoken input and was measured using the Word Error Rate metric (WER). Word Error Rate is the most popular method for rating speech recognition performance. It is calculated by transcribing input utterances and then using the transcriptions as reference to determine the value of recognition results. This metric is based on the Levenshtein distance, which is defined as “the minimum number of editing steps —insertions, deletions and substitutions— needed to match two sequences” (Koehn, 2010, p. 224). The formula for calculating the Word Error Rate is expressed as follows:

\[
\text{WER} = \frac{\text{insertions} + \text{substitutions} + \text{deletions}}{\text{reference word count}}
\]
A couple of tests and a questionnaire were designed to collect the necessary data to measure these attributes and address all other relevant questions. Subjects were presented with a simulation exercise that consisted in administering a household survey about malaria prevention measures to an imaginary French-speaking respondent, and then they were asked to answer a questionnaire about their experience. The following sections provide more details about these specific tasks and other aspects of the evaluation. Section 4.2, in particular, deals with the objectives of the evaluation and specifies the relationship between objectives, attributes and metrics.

4.2 Objectives

The evaluation had three main objectives that were intended to determine the viability of the project:

1. To determine the ability of the system to predict input that differs from the text displayed at the top of the screen in the operative tab of the tool.
2. To determine the ability of the system to recognize spoken input.
3. To determine the ability of the system to translate spoken input.

The term other-than-$PrintName$ input will be used throughout this chapter to refer to any utterance produced by the user that differs from the text displayed at the top of the screen in the operative tab of the tool, which matches a question or a command in a model paper-based questionnaire. For example, the utterance “Do you own a TV?” would be considered as an instance of other-than-$PrintName$ input for the third Group of the main file in the evaluation (“Does your household have a television”). Similarly, the utterance “What is your floor made of?” would count as other-than-$PrintName$ input for the fifth group of the same file (“Main material of the dwelling floor”).

The first objective was worth pursuing because a significant degree of variation in the formulation of questions would support the hypothesis that robustness, i.e., the ability to effectively recognize variant input, is a desirable quality for a speech translation device intended for use in household surveys. Similarly, a significant percentage of other-than-$PrintName$ input that falls within the range of coverage provided by the system would suggest that the language model of the tool might prove suitable to cover the needs of users in the domain at hand.
This objective was set to shed light on one basic generic aspect of household surveys that remained unclear throughout the development phase of the tool—namely, to what extent interviewers would limit themselves to reading the questions out loud exactly as they appear in the paper-based questionnaire or whether they would change the wording of some questions.

There are at least two reasons why an interviewer could change the wording of questions. The first one is to avoid constructions that are atypical of spoken language, since written language tends to be more complex and intricate than speech, with longer sentences and many subordinate clauses. The second reason is that interviewers might need to use their own judgment to adapt the questions depending on the context. For example, if the interviewer can see that there is a bicycle in the room where they are administering the survey, they will probably want to refrain from asking if any member of the household owns a bicycle (the suggested formulation), and ask instead something like “Does that bike over there belong to a member of your household?” or “Whose bike is that?”.

Finally, the second and third objectives of the evaluation were set to estimate the overall performance achieved by the system in the context of the evaluation task.

Each of these objectives correlates to a measurable attribute according to the quality model sketched in Section 4.1. Table 5 below shows the relationship between objectives, attributes and metrics.

<table>
<thead>
<tr>
<th>OBJECTIVES</th>
<th>ATTRIBUTES</th>
<th>METRICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To determine the ability of the system to predict other-than-Printname input</td>
<td>• Suitability (robustness)</td>
<td>• Binary / Annotation</td>
</tr>
<tr>
<td>2. To determine the ability of the system to recognize spoken input</td>
<td>• Accuracy</td>
<td>• Numerical / WER</td>
</tr>
<tr>
<td>3. To determine the ability of the system to translate spoken input</td>
<td>• Suitability (task performance)</td>
<td>• Binary / Annotation</td>
</tr>
</tbody>
</table>

*Table 5. Objectives, attributes and metrics.*

### 4.3 Design

The evaluation included three successive tasks.
The first task was included so that participants could familiarize themselves with the tool before conducting the main experiment. Subjects were asked to play the role of English-speaking interviewers administering a household survey to an imaginary French-speaking respondent. The brief mock survey was made up of three questions: **HC1A.** WHAT IS THE RELIGION OF THE HEAD OF THIS HOUSEHOLD?; **HC1B.** WHAT IS THE MOTHER TONGUE/NATIVE LANGUAGE OF THE HEAD OF THIS HOUSEHOLD?; and **HC1C** WHAT IS THE ETHNIC GROUP OF THE HEAD OF THIS HOUSEHOLD?

All three questions were extracted from the Household Characteristics module of the Multiple Indicator Cluster Survey developed by UNICEF—one of the model paper-based questionnaires that were used to build the corpus (see Section 3.3). These questions were chosen over others because there was a stronger thematic cohesion among them than in any other triplet of consecutive questions in the corpus and because they made up a fairly simple, straightforward task.

The main reason for including a warm-up session was so that participants could check that the system was working properly and that they knew how to use it. One particular concern was that subjects might have had difficulty picking up on the push-and-hold mechanism of the tool, as suggested by previous experiments carried out with similar speech translation tools (Starlander et al., 2005). Accordingly, participants were instructed to keep the input button pressed while speaking their questions, in order to minimize recognition errors caused by truncated input (see Appendix A).

The second task consisted in administering an 18-question sample of the corpus (see Section 3.3). The procedure used to obtain the sample is described in Section 4.3.4 below. This experiment was the primary task of the evaluation and was designed to achieve all three goals described in Section 4.2.

Finally, the third task consisted in answering a brief questionnaire about the potential of the tool in humanitarian settings. Section 4.3.6 provides detailed information about the questionnaire.

**4.3.1 Pilot testing**

The experiments were submitted to a preliminary phase of testing intended to assess a number of aspects of the evaluation design, including the clarity of the instructions and the time required
to complete each task. Pilot testers conducted the warm-up module in about 5 minutes, and the main experiment in about 20 minutes. Both experiments took an average of 20-25 minutes. Adding five more minutes to answer the questionnaire, the estimated time to complete the whole evaluation was set at 30 minutes.

The participants of the pilot test were not part of the target population, nor did they fit the profile of actual participants; most of them were master’s students at the School of Translation of the University of Geneva. Consequently, the piloting was not intended to elicit technical considerations about the deployment of the tool in real contexts.

All out-of-coverage input produced by pilot participants was recorded and incorporated into the system, and a few changes were made to the instructions on the basis of their feedback. Participants were personally supervised during the whole pilot experiment, so the challenges presented by the web-based nature of the actual collection method could not be adequately assessed.

### 4.3.2 Collection Method

The instructions for conducting the test were sent to participants by e-mail in the form of a PDF file on March the 20th 2015. The last data were collected on July the 17th 2015. The file contained a first set of instructions on how to log in to the application and conduct the experiments, and a second set of instructions on how to use it. The latter are available in Appendix A.

Subjects logged in to the application using password-protected accounts and ran the evaluation questionnaire once. The data were collected from a variety of locations, including Afghanistan, Switzerland and Djibouti.

### 4.3.3 Participants

There were three requirements to participate in this study: 1) being employed by a humanitarian organization at the moment of conducting the test; 2) having field experience; and 3) being familiar with household surveys.

We recruited participants with field experience because we were interested in the opinion of humanitarian workers who have real contact with beneficiaries so they could confirm whether
language is indeed a problem for them, and so that we could ask them if they consider household surveys to be an adequate context of use for a speech translation tool.

Being familiar with household surveys could mean anything from having been part of a household survey as an interviewer, to having helped to develop a survey or having studied about them in school. For this reason, the questionnaire that participants answered at the end of the experiments (discussed in Section 4.3.6) included the following question: “How familiar are you with the data collection method presented here, i.e., household surveys?” The purpose of this question was to make sure that all participants perceived themselves as having at least some knowledge about household surveys. Thus, participants who answered “Not at all familiar” would be excluded from the study.

Knowledge about public-health issues was not among the requirements because one of the goals of having the tool tested by humanitarian professionals was to find out whether there were other practical applications within the humanitarian sector that would prove more suitable for an SLT device. However, recruiting efforts were focused on humanitarian professionals who specialize in public-health issues.

The tool was tested by a senior officer from the headquarters of Doctors Without Borders (MSF), one officer from the International Committee of the Red Cross (ICIC), one expert in humanitarian interpreting from the InZone project⁴, and several officers from the United Nations High Commissioner for Refugees (UNHCR). All participants were personally recruited and are identifiable.

4.3.4 Sample

The corpus was sampled so that the time required to complete the main experiment would not exceed a maximum of twenty minutes.

All 63 sentences in the corpus that are capable of eliciting speech by the interviewer⁵ were sorted out according to two criteria: length (sentences that are shorter and longer than 50 characters) and type (yes-no questions, wh questions, disjunctive questions and commands).

⁴ inzone.unige.ch (accessed in August 2015)
⁵ A few sentences in the questionnaires are meant for the interviewer only and, although they are to be displayed at the top of the screen in the tool, they are not intended to elicit speech.
Then, an adequate number of sentences was selected from each sub-group (i.e. long yes-no questions, short yes-no questions, long wh questions, short wh questions, etc.) so that sub-groups in the sample would be relatively proportionate to their corresponding sub-groups in the corpus.

Figure 22 below shows the relation that exists between the sample and the corpus as to sentence length.

![Figure 22. Sampling by sentence length.](image)

Figure 23 shows the relation between the sample and the corpus as to the type of sentences.

![Figure 23. Sampling by type of sentence.](image)
As can be seen in Figure 23, the number of commands in the sample is significantly higher than the number of commands in the corpus. This is due to the fact that commands are more likely to elicit other-than-PrintName input than questions, and a sensible amount of this kind of input was needed to assess the language model of the recognition engine. A complementary way of collecting additional other-than-PrintName input is discussed in Section 4.3.5.

In addition, the selection process was carried out so that the two main subjects of sentences in the corpus, i.e., household characteristics and antimalarial measurers, would be proportionally represented in the sample as well.

Figure 24 shows the relation between the sample and the corpus as to the subject of the sentences.

![Figure 24. Sampling by subject.](image)

Lastly, sentences were grouped according to their corresponding thematic modules and numbered following the order in which they normally appear in the model paper-based questionnaires, i.e., questions about household characteristics in the first place, followed by questions about water and sanitation facilities, then questions about insecticide treated nets and finally questions about indoor residual spraying.
Table 6 below shows all 18 sentences that made the sample distributed according to length and type. The numbers in square brackets signal the sequence structure that was assigned to the evaluation questionnaire. Cells colored in a darker shade of grey display questions about household characteristics; cells colored in a lighter shade of grey contain questions about antimalarial measures.

<table>
<thead>
<tr>
<th>Yes-no</th>
<th>Wh</th>
<th>Command</th>
<th>Disjunctive</th>
</tr>
</thead>
<tbody>
<tr>
<td>LONG</td>
<td>[12] Does your household have any mosquito nets that can be used while sleeping?</td>
<td>[1] How many rooms in this household are used for sleeping?</td>
<td>[6] Is the cooking usually done in the house, in a separate building or outdoors? POP!</td>
</tr>
<tr>
<td></td>
<td>[17] At any time in the past 12 months, has anyone come into your dwelling to spray the interior walls against mosquitoes?</td>
<td>[9] What kind of toilet facility do members of your household usually use?</td>
<td></td>
</tr>
</tbody>
</table>

### 4.3.5 The pop-up message

A pop-up message prompting participants to reformulate their questions was set to be displayed seven times throughout the main experience of the evaluation in order to elicit additional other-than-

-PrintName input.
The message was set to be displayed in one predefined fixed instance from each subgroup in the sample, as shown in Table 6.

![A Prototype Form Filling Application](image)

**Figure 25.** Pop-up message.

The message appeared at the bottom center of the operative tab and read: “Imagine the person you are talking to did not understand your question. Please ask another question that allows you to obtain the information you need.” The phrasing was intended to imply that participants could ask an alternative question with a different meaning, and not just a different version of the same question they had just asked.

The message was set to appear after a first successful recognition so it would not overlap with a “Nothing recognized!” message, which would not have made any sense, since the latter signals that the system has failed to recognize input from the interviewer.

On the one hand, the fact that participants could see the pop-up message only if their utterances were recognized by the system was likely to have a negative impact on the uniformity of the collected data. On the other hand, confronting users with these two scenarios — “machine did not understand” vs. “interlocutor did not understand” — could raise interesting questions about the interactional nature of communication by means of a speech translation device.
4.3.6 The questionnaire

Participants were asked to answer a brief on-line questionnaire about the potential of the tool in humanitarian settings (Appendix B). The questionnaire was hosted on a specialized data-collection website\textsuperscript{16} and contained eight closed questions and two open questions.

The questionnaire included a first group of closed questions that asked participants about their age and level of English, as well as how accustomed they were to exploring new technologies and whether they would be motivated to learn to use a speech translation device. These questions would help sketch a first profile of potential users.

The other closed questions demanded the opinion of participants on whether they were satisfied with the overall performance of the application, how often they encounter language barriers in their jobs and whether the tool might prove useful for its intended context of use.

Finally, two open questions were included so that participants could express their thoughts about other contexts of use that could prove more suitable for this application in the humanitarian sector, and provide comments on how to improve the tool.

4.4 Conclusion

The evaluation of AidSLT was focused on determining the viability of the project and informing its development. One of the central questions was whether interviewers would profit from a certain degree of expressivity, given the characteristics of the domain. The other big goals were to determine the ability of the tool to recognize speech input by English-speaking users and to translate that input into French in the context of the evaluation task.

The evaluation also aimed at a number of subsidiary goals, such as gathering information to start sketching a profile for humanitarian workers as potential users of SLT systems, and determining whether household surveys were a good choice as an application domain for an SLT device.

The necessary data to address these questions was collected by means of two simulation tests and a satisfaction questionnaire. Subjects were presented with an exercise that consisted in

\textsuperscript{16} \url{surveymonke...} (accessed in August 2015)
administering a household survey about malaria prevention measures to an imaginary French-speaking respondent, and a questionnaire about user experience.

The results of the evaluation, as well as the responses to the questionnaire, are presented in detail in the following chapter.
Chapter Five

V. Results

This chapter presents the results of the evaluation in relation to the quality model presented in Section 4.1 and the objectives details in Section 4.2.

Section 5.1 is included as a brief reminder of the objectives that were set for the evaluation, as well as of the specific attributes that were prioritized and the metrics used for measuring them. Section 5.2 presents the data collected through the simulation exercise. Section 5.3 discusses the data collected by means of the questionnaire, including both open and closed questions. Section 5.4 concludes by analyzing the most significant results.

5.1 Introduction

The top-level characteristic prioritized in testing the prototype of AidSLT was the functionality of the tool. The evaluation focused on determining if the tool was likely to perform with certain degree of success in its intended context of use. The two most relevant sub-characteristics, or attributes of the overall functionality of AidSLT were its accuracy and its suitability. Both of these terms are to be interpreted in the light of the ISO standard described in the previous chapter.

Accuracy was associated with the quality of the speech recognition and was to be measured in the most traditional way, i.e., using the Word Error Rate metric. Suitability was the sub-characteristic chosen to describe two separate functionalities of the tool: 1) its ability to predict variant input (robustness) and 2) its ability to translate spoken input in the context of the evaluation task (task performance). These latter two functionalities were measured using binary metrics.

Table 5 in Section 4.2 in the previous Chapter illustrates the relation between objectives, attributes and metrics.
5.2 The simulation exercise

We obtained a total of 221 interactions, of which 134 were correctly logged. One subject misunderstood the instructions, one had severe audio problems with their connection, and a few utterances were spoiled by incorrect use of the push-to-talk interface.

The collected data were annotated independently by two judges. Annotators were asked to transcribe the recorded utterances and answer two questions for each utterance: a) whether the subject appeared to be reading the heading for the questionnaire field or expressing themselves freely, and b) whether the translation produced adequately expressed the question asked in the context of the questionnaire task. The instructions that were given to the annotators are available in Appendix C.

Agreement between the two judges was measured according to two coefficients: Cohen’s Kappa and the Interclass Correlation Coefficient (ICC). The former is the most commonly used in computational linguistics. It is considered to be more robust than simple percent agreement calculation (often referred to as “observed agreement”) because it takes into account the hypothetical probability of chance agreement between the two raters (Gerlach, 2015). The scores obtained for Cohen’s Kappa were very good: 0.786.

The Interclass Correlation Coefficient was computed as a complementary measure. This coefficient presents the advantage of taking into account the magnitude of disagreement between raters, and it is considerably less used in the field (Gerlach, 2015). The scores obtained for the ICC (0.922) indicated excellent agreement between the two judges.

Between 73 and 75 of the interactions were judged to be freely expressed (54%-56%), while 59 and 61 were judged to be read (44%-46%). Contrary to what was expected, the majority of the input was freely expressed. The evaluation questionnaire was made up of 14 questions and 4 commands. It was possible that interviewers would limit themselves to reading the questions out loud and producing freely-expressed phrases only when the header was a command (e.g. “Ask permission to observe the facility”), so we prompted them to rephrase their questions seven times throughout the exercise to collect additional other-than-PrintName input (see Section 4.3.5). Still, the number of interactions elicited by the 7 pop-up messages and the 4 commands was lower than the interactions elicited by questions, which means that subjects
changed the wording of questions the first time they said them, without being prompted to do so.

There were 70 interactions that were judged to be freely expressed by both annotators. Out of these interactions, 26 (37%) were covered by the system, i.e., they could be generated by the grammar exactly as they were produced by the participants of the evaluation, and 44 (63%) fell outside the coverage of the tool. Exactly half of those 44 out-of-coverage utterances were just two editing steps (insertions, deletions and substitutions) away from the closest variant available in the grammar. This means that the domain expert who defined the content managed to predict 37% of all other-than-Printname input.

Figure 26 below shows the proportion of in-coverage and out-of coverage input for utterances that were judged to be freely expressed.

![Figure 26. Freely expressed interactions.](image)

All 26 freely-expressed in-coverage utterances were successfully recognized and produced correct translations, as did all utterances judged to be read. More significantly, the translated versions of the utterances that were freely expressed and not covered by the system were correct in most cases. One of the annotators rated 10 translations as incorrect, while the other one did the same for only 6 (out of 44). Even considering the least indulgent annotations, the proportion of adequate translations for out-of-coverage input was very high (77%). Naturally, the proportion of correct translations over the whole set of analyzed interactions was comparatively higher (93%).
All incorrect translations were due to incorrect recognition results produced by out-of-coverage input.

The system associates input utterances to the available variant (in the corresponding Group) that is acoustically most similar. As illustrated by the figures above, the translation strategy deployed in AidSLT reacted very well to out-of-coverage input, since the vast majority of utterances that were not covered by the system were adequately translated. For example, the recognition result for “Where do you go fetch your water?”, an input utterance produced by one of the subjects, was “Where do you get your water?”. Similarly, the input utterance “Where is the cooking usually done?” was recognized as “Where do you usually cook?”.

However, there were a few cases where the most acoustically similar variant in the Group did not convey the meaning of the input utterance. For example, the question “How many members of this household own a bicycle?” was recognized as “Does any member of this household own a bike?”. Similarly, “Who uses this toilet?” was recognized as “May I please see the toilet?”. This highlights the importance of displaying the recognition result on the screen so that the user can approve or disapprove of it before eliciting the translation.

The Word Error Rate, as expected, was rather high as well (29%). The fact that the Word Error Rate was high and the proportion of incorrect translations over the total of input utterances was very low (about 7%) suggests that the tool exploits the constrained nature of the task well.
5.3 The questionnaire

After completing the simulation exercise, participants answered a brief on-line questionnaire intended to sketch a first profile of humanitarian workers as potential users of rule-based SLT systems and to determine the potential of the tool in humanitarian settings.

We recruited participants with field experience because we were interested in the opinion of humanitarian workers who have real contact with beneficiaries so they could confirm that language is indeed a problem for them, and we could ask them whether they consider household surveys to be an adequate context of use for a speech translation tool.

Participants were presented with 8 multiple-choice questions and two open questions. The latter asked them to suggest other possible application domains for AidSLT and provide comments on how to improve the tool. The design of the questionnaire can be consulted in Appendix B.

5.3.1 Closed questions

All subjects declared themselves to be proficient or advanced speakers of English (although only one of them spoke English as their first language) and to be moderately familiar or very familiar with household surveys. The purpose of these questions was to make sure that all subjects were qualified to take the test, namely, that they could interact with the system in English and that they could give meaningful opinions about the intended context of use.

Five subjects said they were between 25 and 34 years old, one was between 35 and 44, and one between 45 and 54. Most subjects said they were accustomed to being presented with new kinds of technologies (5 to 2) and they were willing to invest time in learning to use a speech translation tool (6 to 1).

The question about how satisfied they were with the overall performance of the tool was the one that presented the most disparity. Two subjects said they were quite satisfied, two said they were somewhat satisfied, two said they were neither satisfied nor dissatisfied, and one declared themselves to be somewhat dissatisfied. The only participant who declared themselves to be dissatisfied with the performance of the tool had misunderstood the instructions, attempting to answer the header questions instead of asking them; it is reasonable that the experience for this user was far from satisfying.
Finally, all participants perceived language barriers to be very often encountered in humanitarian settings and claimed that AidSLT might prove useful for bridging language barriers between interviewers and respondents in the context of a household survey. These are two of the most significant data concerning the viability of the project.

5.3.2 Open questions

The two last questions of the questionnaire asked participants to express their thoughts about other contexts of use that could prove more suitable for this application in the humanitarian sector and to provide comments on how to improve the tool.

All participants suggested numerous alternative applications for AidSLT, except for one, who said that the most suitable context of use was household surveys. Some of the suggestions included needs assessment exercises, basic conversations with refugees in the context of interviews for registration, resettlement and status determination, and data collection focused on ebola. One participant suggested that a more robust version of AidSLT could be deployed in meetings with beneficiaries in focus groups interviews, or in meetings with other counterparts, such as governments or local NGOs.

Likewise, participants offered several sensible comments on how to improve the tool. Some subjects reported specific problems related to the need of having Adobe Flash Player installed in their computers and the microphone activation mechanism. Some suggestions pointed simply at developing a better, more efficient tool: to make it bidirectional, to add more languages, to make it a phone app, to make it faster. One sensible comment about the interface was to change the layout of the answers tab to include a higher number of options.

One participant observed that countries in need of humanitarian assistance typically have unreliable internet connectivity, and that therefore it would be better if the tool functioned offline. This participant also proposed that the tool support offline extraction of data in an adequate analysis format, such as csv or xls.

One participant showed interest in the ability of the tool to "correct" the formulation of the question. This participant said that it is often the case that some interviewers use words or phrases that might be outdated or that might seem offensive to respondents, such as the term "deaf-and-dumb" to refer to people with a hearing impediment. Because the formalism of the tool matches a number of questions with one unique translation, the developer of the survey
could "control" to some extent the language of the interviewer. Also, if we included the back translation of the utterance in the tab for the interviewer, they could see their mistakes and learn from them. So in this sense the tool could also be educational for the interviewer.

5.4 Conclusion

The percentage of utterances that were freely expressed by subjects (54%–56%) was very high, suggesting that interviewers are likely to change the wording of header questions and that robustness is indeed a desirable quality for AidSLT. Similarly, the high percentage of predicted input (37%) suggests that the formalism seems suitable to cover the expressive needs of users in the domain at hand. These two conclusions are related to the first objective of the evaluation.

Regarding the second objective, if we analyze the ability of the tool to recognize spoken input in isolation, the results are not that good; 29% is a rather high figure for Word Error Rate. However, this figure should be analyzed together with the Task Error Rate (the ability of the tool to translate spoken input in the context of the questionnaire task), which was very low (7%).

The combination of a high Word Error Rate and a low Task Error Rate suggests that the tool exploits the constrained nature of the domain well. In-coverage input was almost always well recognized and well translated. Similarly, out-of-coverage input was mostly directed towards semantically adequate variants. Still, a comparatively low percentage of incorrect translations points to the need to display the recognition result on screen for verification purposes.

The most significant data collected through the questionnaire are listed below.

1. Subjects are accustomed to being presented with new kinds of technologies and willing to invest time in learning to use a speech translation tool; this suggests that humanitarian workers would make good users of the tool.
2. Participants perceive language barriers as a major problem in humanitarian settings; they think AidSLT might prove useful for its intended context of use and that there are other domains where it might be applied.
3. Participants had problems opening the application due to the Adobe Flash Player requirements and said that AidSLT should be able to function off-line, since it is intended to be deployed in regions of the world with limited access to the internet.
4. Participants also had difficulty picking up on the push-to-talk mechanism and suggested that we change the layout of the PROVIDE ANSWERS tab to include a higher number of options.

5. An advantage of the tool (that we were unaware of before testing the prototype) is that domain experts defining content can use the tool to correct the language of interviewers and to teach them how to phrase questions in an appropriate manner.
VI. Conclusions and Directions for Future Work

The results of the evaluation are very encouraging. The most significant conclusion is that the overall architectural approach seems adequate for the targeted domain. The language model and the translation strategy favor a healthy balance between robustness and accuracy: the tool is likely to respond well to the expressive needs of users in the context of the application domain and to produce correct translations of input utterances.

Further development should be centered on two main aspects. One is the development of content by domain experts. The level of performance of the tool depends largely on how relevant the content is and how thoroughly it is defined. The first step in this direction must be to analyze and describe in full the application domain. The ideal way to do this would be to witness real-life interactions between humanitarian workers and beneficiaries in the context of household surveys.

Several attempts were made at the onset of the project to engage humanitarian organizations in the development of the prototype, in the hope that a conjoint approach would result in a phase of field testing. However, these efforts proved unsuccessful because development was at a very early stage. The results of this evaluation can be used as a basis to resume those efforts.

The second aspect is also concerned with content development, but from a linguistic engineering perspective. It should be interesting to explore the extent to which regular expression grammars can be used to support the creation of nontrivial domain coverage and to address translation problems like the ones described in Section 2.1.
References


Bouillon, P., Clas, A. & UREF, 1993. La Traductique: études et recherches de traduction par ordinateur


Koehn, P., 2010. Statistical Machine Translation,

Mariani, J., 2010. Spoken Language Processing.


Rayner, E., Bouillon, P. & Hockey, B.A., 2006. Putting Linguistics into Speech Recognition,


72
Appendix A

HOW TO USE THE APP

The **provide answer** tab displays a human translation of the recognition result along with a set of relevant predefined answers. This tool can be deployed in delicate contexts — such as public health surveys or doctor-patient dialogues — because users can verify that the system has “understood” them correctly before delivering their messages, and because translations are highly reliable.

**Recognition result.** This is what the machine “understood” of what you said while you were holding the input button.

**Question as it appears in the corresponding paper-based questionnaire.**

Most of the questions include a set of instructions that will help you understand the exact purpose of what you are asking.

**Keep hold of the mic button while speaking your question!**

Always make sure that:

1. The background of the icon is orange when you are speaking to the system.
2. You can hear your voice back right after having released the button.

Press the Help button to get a list of example questions.

Click on any answer to move on to the next question. Use the arrows to navigate through the survey.
1. What is your age?
   - 18 to 24
   - 25 to 34
   - 35 to 44
   - 45 to 54
   - 55 to 64
   - 65 to 74
   - 75 or older

2. What would you say is your level of English?
   - Basic
   - Intermediate
   - Advanced
   - Proficient

3. Are you accustomed to being presented with new kinds of technologies?
   - Yes
   - No

4. Would you be willing to invest time in learning to use a speech translation tool?
   - Yes
   - No

5. How familiar are you with the data collection method presented here, i.e., household surveys?
   - Extremely familiar
   - Very familiar
   - Moderately familiar
   - Slightly familiar
   - Not at all familiar
6. Overall, are you satisfied, dissatisfied, or neither satisfied nor dissatisfied with the performance of the speech recognition?
   - Extremely satisfied
   - Quite satisfied
   - Somewhat satisfied
   - Neither satisfied nor dissatisfied
   - Somewhat dissatisfied
   - Quite dissatisfied
   - Extremely dissatisfied

7. Which of the following statements do you agree with most strongly?
   - Language barriers are very often encountered by humanitarian workers.
   - Language barriers are often encountered by humanitarian workers.
   - Language barriers are sometimes encountered by humanitarian workers.
   - Language barriers are rarely encountered by humanitarian workers.
   - Language barriers are never encountered by humanitarian workers.

8. Do you think this tool might prove useful for its intended purpose, i.e., bridging language barriers between interviewers and respondents in the context of a household survey?
   - Yes
   - No

9. Can you think of any other contexts of use that might prove more suitable for a speech translation tool in humanitarian settings?

10. How would you improve this tool?
Appendix C

Instructions for annotating the questionnaire data

1. The spreadsheet shows the recognition result and the translation for each question. The audio link allows the annotator to listen to the recorded speech.

2. For each line, the annotator should listen to the recording and edit the text under “correct words” to provide a correct transcription.

3. The annotator should fill in the field “Judgment” to reflect two possibilities:
   
   a. Either the subject read the header question, or they formulated themselves freely.

   b. Either the translation produced was adequate in the context of the questionnaire task, or it was not.

The annotator marks these as Read/Correct, Read/Incorrect, Free/Correct or Free/Incorrect.
Appendix D

GROUPS

Group
Questionnaire EVALUATION
Name RoomsUsedForSleeping
PrintName How many rooms in this household are used for sleeping?
Code 1
HelpFile 1.xml
Fillers one two three four
Next Refrigerator
EndGroup

Group
Questionnaire EVALUATION
Name Refrigerator
PrintName Does your household have a refrigerator?
Code 2
HelpFile 2.xml
Fillers yes no
Next Television
EndGroup

Group
Questionnaire EVALUATION
Name Television
PrintName Does your household have a television?
Code 3
HelpFile 3.xml
Fillers yes no
Popup The respondent did not understand. Please ask another question that allows you to obtain the information you need.
Next Bicycle
EndGroup

Group
Questionnaire EVALUATION
Name Bicycle
PrintName Does any member of this household own a bicycle?
Code 4
HelpFile 4.xml
Fillers yes no
Next MaterialOfFloor
EndGroup

Group
Questionnaire EVALUATION
Name MaterialOfFloor
PrintName Main material of the dwelling floor.
Code 5
HelpFile 5.xml
Fillers earth/sand dung wood_planks palm/bamboo parquet_or_polished_wood vinyl_or_asphalt_strips ceramic_tiles cement carpet other
Popup The respondent did not understand. Please ask another question that allows you to obtain the information you need.
Next CookingPlace
EndGroup
Group Questionnaire EVALUATION
Name CookingPlace
PrintName Is the cooking usually done in the house, in a separate building or outdoors?
Code 6
Fillers in the house in a separate building outdoors
Popup The respondent did not understand. Please ask another question that allows you to obtain the information you need.
Next WaterForDrinking
EndGroup

Group Questionnaire EVALUATION
Name WaterForDrinking
Code 7
HelpFile 7.xml
PrintName What is the main source of drinking water for members of your household?
Fillers piped into dwelling piped to yard/plot piped to neighbour public tap/standpipe tube well or borehole protected well unprotected well protected spring unprotected spring rainwater collection tanker truck cart with small tank surface water bottled water other
Popup The respondent did not understand. Please ask another question that allows you to obtain the information you need.
Next SourceLocation
EndGroup

Group Questionnaire EVALUATION
Name SourceLocation
PrintName Where is that water source located?
Code 8
Fillers in own dwelling in own yard/plot elsewhere
Next KindOfToilet
EndGroup

Group Questionnaire EVALUATION
Name KindOfToilet
PrintName What kind of toilet facility do members of your household usually use?
Code 9
HelpFile 9.xml
Fillers flush/pour flush ventilated improved pit latrine pit latrine with slab pit latrine without slab/open pit composting toilet bucket hanging toilet hanging latrine no facility/bush/field other
Next PermissionToObserve
EndGroup

Group Questionnaire EVALUATION
Name PermissionToObserve
PrintName Ask permission to observe the facility.
Code 10
Fillers yes no
Next PublicFacility
EndGroup

Group
Questionnaire EVALUATION
Name PublicFacility
PrintName Do you share this facility only with members of other households that you know, or is the facility open to the use of the general public?
Code 11
HelpFile 11.xml
Fillers other_households_only public_facility
Next NetsWhileSleeping
EndGroup

Group
Questionnaire EVALUATION
Name NetsWhileSleeping
Code 12
HelpFile 12.xml
PrintName Does your household have any mosquito nets that can be used while sleeping?
Fillers yes
Next ShowNets
EndGroup

Group
Questionnaire EVALUATION
Name ShowNets
PrintName Ask the respondent to show you the nets in the household.
Code 13
HelpFile 13.xml
Fillers yes no
Popup The respondent did not understand. Please ask another question that allows you to obtain the information you need.
Next BrandOfNet
EndGroup

Group
Questionnaire EVALUATION
Name BrandOfNet
PrintName Observe or ask the brand/type of mosquito net. If brand is unknown and you cannot observe the net, show pictures of typical net types/brands to the respondent.
Code 14
HelpFile 14.xml
Fillers brand_A brand_B brand_C dk_brand brand_D brand_E brand_F other
Next AnyoneSleptUnder
EndGroup

Group
Questionnaire EVALUATION
Name AnyoneSleptUnder
PrintName Did anyone sleep under this mosquito net last night?
Code 15
Fillers yes
Popup The respondent did not understand. Please ask another question that allows you to obtain the information you need.
Next WhoSleptUnder
EndGroup

Group
Questionnaire EVALUATION
Name WhoSleptUnder
PrintName Who slept under this net last night?
Code 16
HelpFile 16.xml
Fillers name1 name2 name3
Next WallsSprayed
EndGroup

Group
Questionnaire EVALUATION
Name WallsSprayed
PrintName At any time in the past 12 months, has anyone come into your dwelling to spray the interior walls against mosquitoes?
Code 17
HelpFile 17.xml
Fillers yes
Next WhoSprayed
EndGroup

Group
Questionnaire EVALUATION
Name WhoSprayed
PrintName Who sprayed the dwelling?
Code 18
Fillers government_worker/program private_company ngo other dk/not_sure
Popup The respondent did not understand. Please ask another question that allows you to obtain the information you need.
Next Thanks
EndGroup

Group
Questionnaire EVALUATION
Name Thanks
PrintName Thanks for having helped us evaluate this tool.
Code 18
Fillers thanks
Next Thanks
EndGroup

QUESTIONS

# Questions for 'RoomsUsedForSleeping'

Question
Group RoomsUsedForSleeping
Translation/French Combien de pièces utilisez-vous pour dormir dans ce ménage ?
Translation/Spanish ¿Cuántos cuartos de esta vivienda se utilizan para dormir?
Variant how many rooms ?( here | in this household ) are used for sleeping ?( on a regular basis ) ?( here | in this household )
Variant how many rooms do you use for sleeping ?( in this household ) ?( on a regular basis )
Variant how many rooms ?( in this household ) are used for sleeping
Variant ?( in ) how many rooms do you sleep ?( in )
Answers one two three four
EndQuestion

Question
Group RoomsUsedForSleeping
Translation/French Est-ce que quelqu'un utilise cette salle-ci pour dormir ?
Translation/Spanish ¿Alguien usa este cuarto para dormir?
Variant ( do | does ) ( people | anyone | any person ) ?( usually ) sleep in ( this | these | that | those ) ( room | rooms ) ?( here ) ?( over ) ?( there ) ?( on a regular basis )
Answers yes no
EndQuestion

# Questions for 'MaterialOfFloor'

Question
Group MaterialOfFloor
Translation/French Quel est le principal matériau du sol ?
Translation/Spanish ¿Cuál es el material principal del piso?
Variant what is the ?( main ) material of the ?( dwelling ) floor
Variant what is the ?( main ) type of ( floor | flooring ) ?( in ) ?( this | the | your ) ?( house | household | dwelling )
Variant ( which | what ) material covers the largest amount of floor ?( space ) ?( in ) ?( this | the | your ) ?( house | household | dwelling )
Variant what ?( material ) is the floor ?( principally | mainly | mostly ) made of
Answers earth/sand dung wood_planks palm/bamboo parquet_or_polished_wood vinyl_or_asphalt_strips ceramic_tiles cement carpet other
EndQuestion

QuestionTemplate yes_no_floormaterial FRENCH SPANISH ENGLISH VALUE
Group MaterialOfFloor
Translation/French Le sol du ménage est-il principalement fait de FRENCH ?
Translation/Spanish ¿El piso de su casa está hecho principalmente de SPANISH?
Variant is ( the | your ) floor ?( in your house ) ?( principally | mainly | mostly ) made of ENGLISH
Answers yes=VALUE no
EndQuestionTemplate

ApplyTemplate yes_no_floormaterial "sable" "arena" "sand" "earth/sand"
ApplyTemplate yes_no_floormaterial "terre" "tierra" "earth" "earth/sand"
ApplyTemplate yes_no_floormaterial "bouse" "estiercol" "dung" "dung"
ApplyTemplate yes_no_floormaterial "planches en bois" "tablas de madera" "wood_planks" "wood_planks"
ApplyTemplate yes_no_floormaterial "palmes" "hojas de palmera" "palm" "palm/boamboo"
ApplyTemplate yes_no_floormaterial "bambou" "bambú" "bamboo" "palm/bamboo"
ApplyTemplate yes_no_floormaterial "parquet" "parquet" "parquet" "parquet_or_polished_wood"
ApplyTemplate yes_no_floormaterial "bois ciré" "madera lustrada" "polished wood" "parquet_or_polished_wood"
ApplyTemplate yes_no_floormaterial "vinyle" "vinilo" "vynil" "vinyl_or_asphalt_strips"
ApplyTemplate yes_no_floormaterial "asphalt" "asfalto" "asphalt | strips | vinyl_or_asphalt_strips"
ApplyTemplate yes_no_floormaterial "carrelage" "cerámicos" "ceramic_tiles"
ApplyTemplate yes_no_floormaterial "ciment" "cemento" "cement" "cement"

Question
Group MaterialOfFloor
Translation/French La plupart du sol est-elle couvert par une moquette ?
Translation/Spanish ¿La mayor parte del piso de su casa está cubierto por una alfombra?
Variant is ( the | your ) floor ?( in your house ) ?( principally | mainly | mostly ) covered by ?( a ) carpet
# Questions for 'Television'

**Question**
Group Television
Translation/French Dans votre ménage, avez-vous la télévision ?
Translation/Spanish ¿Tiene su hogar un televisor?

Variants:
- Does your household have a television?
- Do you own a television?
- Is there a television in your household?
- Are there any televisions in your household?
- What about a television in your household?

Answers: yes no

EndQuestion

# Questions for 'Refrigerator'

**Question**
Group Refrigerator
Translation/French Dans votre ménage, avez-vous un réfrigérateur ?
Translation/Spanish ¿Tiene su hogar un refrigerador?

Variants:
- Does your household have a refrigerator?
- Do you own a refrigerator?
- Is there a refrigerator in your household?
- Are there any refrigerators in your household?
- What about a refrigerator in your household?

Answers: yes no

EndQuestion
# Questions for 'Bicycle'

**Question**

Group Bicycle

Translation/French Est-ce qu'un membre de votre ménage possède une bicyclette ?

Translation/Spanish ¿Algún miembro de su hogar tiene bicicleta?  

Variant does any member of (this | your) (household | house | dwelling) (own | have) a (bicycle | bike)

Variant does anyone in (this | your) (household | house | dwelling) (own | have) a (bicycle | bike)

Variant is anyone in (this | your) (household | house | dwelling) the owner of a (bicycle | bike)

Variant do you (own | have) a (bicycle | bike)

Answers yes no

EndQuestion

# Questions for 'CookingPlace'

**Question**

Group CookingPlace

Translation/French Où est la cuisine habituellement faite ?

Translation/Spanish ¿Dónde se cocina por lo general?

Variant where do you ?(usually | normally | generally | mostly) (cook | do the cooking)

Variant is the cooking ?(usually | normally | generally | mostly) done (in the house | in a separate building | outdoors) ?(or) (in the house | in a separate building | outdoors) ?(or) ?(in the house | in a separate building | outdoors)

Variant do you ?(usually | normally | generally | mostly) (cook | do the cooking) (in the house | in a separate building | outdoors) ?(or) ?(in the house | in a separate building | outdoors)

Answers in_the_house in_a_separate_building outdoors

EndQuestion

**QuestionTemplate yes_no_cookingplace**

**FRENCH** SPANISH ENGLISH VALUE

Group CookingPlace

Translation/French La cuisine est-elle habituellement faite FRENCH ?

Translation/Spanish ¿Se cocina por lo general SPANISH?

Variant is the cooking ?(usually | normally | generally | mostly) done ENGLISH

Variant do you ?(usually | normally | generally | mostly) (cook | do the cooking) ENGLISH

Variant ENGLISH

Answers yes=VALUE

EndQuestionTemplate

ApplyTemplate yes_no_cookingplace "dans la maison" "en la casa" "(in | inside) the house" "in_the_house"

ApplyTemplate yes_no_cookingplace "dans un autre bâtiment" "en otro edificio" "in ?(a) (separate | different | another) building" "in_a_separate_building"

ApplyTemplate yes_no_cookingplace "en plain air" "al aire libre" "outdoors" "outdoors"

# Questions for 'WaterForDrinking'
Question
Group WaterForDrinking
Translation/French D'où provient principalement l'eau que boivent les membres de votre ménage ?
Translation/Spanish ¿Cuál es la fuente principal de agua que beben los miembros de su hogar?
Variant what is the ( main | principal ) source of drinking water ?( for members of your household | in this house )
Variant what is the ( household | dwelling ) ( main | principal ) source of drinking water
Variant ?( from ) where do you ?( principally | mainly | mostly | usually ) get ( your | the ) ?( drinking ) water ?( for drinking | that you drink ) ?( from )
Variant ?( from ) where does ( your | the ) ?( drinking ) water ?( for drinking | that you drink ) ?( principally | mainly | mostly | usually ) come ?( from )
Answers piped_into_dwelling piped_to_yard/plot public_tap/standpipe piped_to_neighbour tube_well_or_borehole protected_well unprotected_well protected_spring unprotected_spring rainwater_collection tanker_truck cart_with_small_tank surface_water bottled_water other
EndQuestion

QuestionTemplate yes_no_drinkingwater FRENCH SPANISH ENGLISH VALUE
Group WaterForDrinking
Translation/French L'eau que boivent les membres de votre ménage vient-elle ?
Translation/Spanish ¿El agua que beben los miembros de su hogar viene ?
Variant do you ?( principally | mainly | mostly | usually ) ( get | collect ) ( your | the ) ?( drinking ) water ?( that you drink ) ENGLISH
Variant does ( your | the ) ?( drinking ) water ?( that you drink ) ?( principally | mainly | mostly | usually ) come ENGLISH
Answers yes=VALUE no
EndQuestionTemplate

ApplyTemplate yes_no_drinkingwater "d'un robinet dans la maison" "de un grifo en la casa" "from a tap ( in | inside ) ( the house | the kitchen | the bathroom )" "piped_into_dwelling"
ApplyTemplate yes_no_drinkingwater "d'un robinet dans le jardin" "de un grifo en el patio" "from a tap ?( outdoors ) ?( in the yard | in the plot | in the compound )" "piped_to_yard/plot"
ApplyTemplate yes_no_drinkingwater "d'un robinet public" "de un grifo público" "from a ( public tap | standpipe )" "public_tap/standpipe"
ApplyTemplate yes_no_drinkingwater "d'un robinet d'un voisin" "de un grifo del vecino" "from the neighbours ( dwelling | house | yard | plot | compound )" "piped_to_neighbour"
ApplyTemplate yes_no_drinkingwater "d'un trou" "de un pozo" "from a ( tube well | borehole )" "tube_well_or_borehole"
ApplyTemplate yes_no_drinkingwater "d'un trou récouvert" "de un aljibe recubierto" "from a ?protected well" "protected_well"
ApplyTemplate yes_no_drinkingwater "d'un trou découvert" "de un aljibe descubierto" "from ( a | an ) ?unprotected well" "unprotected_well"
ApplyTemplate yes_no_drinkingwater "d'une source réouverte" "de un manantial recubierto" "from a protected spring" "protected_spring"
ApplyTemplate yes_no_drinkingwater "d'une source découverte" "de un manantial descubierto" "from ( a | an ) ?unprotected spring" "unprotected_spring"
ApplyTemplate yes_no_drinkingwater "de la pluie" "de la lluvia" "from rain" "rainwater_collection"
ApplyTemplate yes_no_drinkingwater "d'un camion-citerne" "de un camión cisterna" "from a ?( tanker ) truck ?( with a tank )" "tanker_truck"
ApplyTemplate yes_no_drinkingwater "d'une charrette avec une petite citerne" "de una carreta con cisterna" "from a cart with a small tank" "cart_with_small_tank"
ApplyTemplate yes_no_drinkingwater "d'une rivière" "de un río" "from a river" "surface_water"
ApplyTemplate yes_no_drinkingwater "d'un fleuve" "de un arroyo" "from a stream" "surface_water"
ApplyTemplate yes_no_drinkingwater "d'un barrage" "de una represa" "from a dam" "surface_water"
ApplyTemplate yes_no_drinkingwater "d'un lac" "de un lago" "from a lake" "surface_water"
ApplyTemplate yes_no_drinkingwater "d'une mare" "de una laguna" "from a pond" "surface_water"
ApplyTemplate yes_no_drinkingwater "d'un canal" "de un canal" "from a channel" "surface_water"
ApplyTemplate yes_no_drinkingwater "d'un canal d'irrigation" "de un canal de irrigación" "from an irrigation channel" "surface_water"

Question
Group WaterForDrinking
Translation/French Quel genre de trou est-il ?
Translation/Spanish ¿Qué tipo de pozo es?
Variant what ( kind | type | sort ) of well ( would that be | is it )
Answers protected_well unprotected_well
EndQuestion

Question
Group WaterForDrinking
Translation/French Quel genre de source est-elle ?
Translation/Spanish ¿Qué tipo de manantial es?
Variant what ( kind | type | sort ) of spring ( would that be | is it )
Answers protected_spring unprotected_spring
EndQuestion

QuestionTemplate yes_no_protected FRENCH SPANISH ENGLISH VALUE
Group WaterForDrinking
Translation/French S'agit-il d'un FRENCH ?
Translation/Spanish ¿Es SPANISH?
Variant ( would that be | is it ) ENGLISH
Answers yes=VALUE no
EndQuestionTemplate

ApplyTemplate yes_no_protected "un trou récouvert" "un pozo recubierto" "a protected well" "protected_well"
ApplyTemplate yes_no_protected "une source récouverte" "un manantial recubierto" "a protected spring" "protected_spring"

QuestionTemplate yes_no_unprotected FRENCH SPANISH ENGLISH VALUE
Group WaterForDrinking
Translation/French S'agit-il d'un FRENCH ?
Translation/Spanish ¿Es SPANISH?
Variant ( would that be | is it ) ENGLISH
Answers yes=VALUE no
EndQuestionTemplate

ApplyTemplate yes_no_unprotected "un trou découvert" "un pozo descubierto" "an unprotected well" "unprotected_well"
ApplyTemplate yes_no_unprotected "une source découverte" "un manantial descubierto" "an unprotected spring" "unprotected_spring"

QuestionTemplate yes_no_delivery FRENCH SPANISH ENGLISH VALUE
Translation/French L'eau que boivent les membres de votre ménage est-elle livrée chez vous FRENCH?
Translation/Spanish ¿El agua que beben los miembros de su hogar viene ESPANISH?
Variant is ( your | the ) ?( drinking ) water ?( that you drink ) delivered to ( you | your home | your house | your dwelling ) ENGLISH
Answers yes=VALUE no
EndQuestionTemplate

ApplyTemplate yes_no_delivery "dans un camion-citerne" "de un camión cisterna" "in a ?( tanker ) truck ?( with a tank )" "tanker_truck"
ApplyTemplate yes_no_delivery "dans une charrette avec une petite citerne" "de una carreta con cisterna" "from a cart with a ?( small ) ?( tank | drum )" "cart_with_small_tank"

QuestionTemplate yes_no_adj_water FRENCH SPANISH ENGLISH VALUE
Group WaterForDrinking
Translation/French L'eau que boivent les membres de votre ménage est-elle FRENCH?
Translation/Spanish ¿El agua que beben los miembros de su hogar es SPANISH?
Variant is ( your | the ) ?( drinking ) water ?( that you drink ) ENGLISH
Variant do you ?( principally | mainly | mostly | usually ) drink ENGLISH
Answers yes=VALUE no
EndQuestionTemplate

ApplyTemplate yes_no_adj_water "de l'eau en bouteille" "agua envasada" "bottled water" "bottled_water"
ApplyTemplate yes_no_adj_water "de l'eau de pluie" "agua de lluvia" "rainwater collection" "rainwater_collection"

# Questions for 'SourceLocation'

Question
Group SourceLocation
Translation/French Où cette source d'approvisionnement en eau est-elle située ?
Translation/Spanish ¿Dónde se encuentra esa fuente de agua?
Variant where is ( that | the ) ?( water ) source ?( of water ) located
Variant what is the location of ( that | the ) ?( water ) source ?( of water )
Answers in_own_dwelling in_own_yard/plot elsewhere
EndQuestion

QuestionTemplate yes_no_source_location FRENCH SPANISH ENGLISH VALUE
Group SourceLocation
Translation/French Cette source d'approvisionnement en eau est-elle située FRENCH?
Translation/Spanish ¿Esa fuente de agua está SPANISH?
Variant is that ?( water ) source located ENGLISH
Answers yes=VALUE no
EndQuestionTemplate

ApplyTemplate yes_no_source_location "dans votre maison" "en su casa" "in your ?( own ) dwelling" "in your own dwelling"
ApplyTemplate yes_no_source_location "dans votre jardin" "en su patio" "in your ?( own ) yard" "in your own yard/plot"
ApplyTemplate yes_no_source_location "dans votre parcelle" "en su parcela" "in your ?( own ) plot" "in your own yard/plot"
ApplyTemplate yes_no_source_location "ailleurs" "en otra parte" "( somewhere else | elsewhere )" "elsewhere"
# Questions for 'KindOfToilet'

**Question**

Group KindOfToilet

Translation/French Habituellement, quel type de toilettes les membres de votre ménage utilisent-ils ?
Translation/Spanish ¿Qué tipo de instalación sanitaria utilizan por lo general los miembros de su hogar?

Variant what (kind | type) of (sanitation | toilet) do members of your household (usually) use

Variant what (kind | type) of (sanitation | toilet) do you (usually) use

Answers flush/pour_flush ventilated_improved_pit_latrine pit_latrine_with_slab pit_latrine_without_slab/open_pit composting_toilet bucket hanging_toilet_hanging_latrine no_facility/bush/field other

EndQuestion

**QuestionTemplate** yes_no_toilet FRENCH SPANISH ENGLISH VALUE

Group KindOfToilet

Translation/French Est-ce que vous utilisez FRENCH comme toilettes ?
Translation/Spanish ¿La instalación sanitaria que utiliza normalmente en el hogar es SPANISH?

Variant do you (usually) use ENGLISH as a (sanitation | toilet) ?

Variant is your sanitation facility ENGLISH

Answers yes=VALUE no

EndQuestionTemplate

ApplyTemplate yes_no_toilet "des toilettes avec une chasse d'eau" "un inodoro con cisterna" "a flush toilet" "flush/pour_flush"

ApplyTemplate yes_no_toilet "un seau" "un balde" "a bucket" "bucket"

ApplyTemplate yes_no_toilet "des toilettes à compostage" "un inodoro de compostaje" "a composting toilet" "composting_toilet"

# Questions for 'PermissionToObserve'

**Question**

Group PermissionToObserve

Translation/French Est-ce que je pourrais voir les toilettes ?
Translation/Spanish ¿Comparte usted esta instalación únicamente con miembros de otros hogares que usted conoce, o la instalación está abierta al uso del público en general?

Variant do you share this facility only with members of other households that you know or is the facility open to the use of the general public?

Answers yes no

EndQuestion

# Questions for 'PublicFacility'

**Question**

Group PublicFacility

Translation/French Partagez-vous ces toilettes seulement avec des membres d'autres ménages que vous connaissez, ou est-ce que n'importe qui peut utiliser ces toilettes ?

Translation/Spanish ¿Comparte usted esta instalación únicamente con miembros de otros hogares que usted conoce, o la instalación está abierta al uso del público en general?

Variant do you share this facility only with members of other households that you know or is the facility open to the use of the general public?
Variant do you share this facility?( only ) with ( people | members ) ?( of other households ) that you know
Variant is this facility open to the use of the general public
Variant ?( with ) whom do you share the facility ?( with )
Variant is this a public facility
Answers other_households_only public_facility

EndQuestion

# Questions for 'NetsWhileSleeping'

Question
Group NetsWhileSleeping
Translation/French Est-ce que votre ménage possède des moustiquaires qui peuvent être utilisés pour dormir ?
Translation/Spanish ¿Hay en su hogar algún mosquitero que se pueda utilizar para dormir?
Variant does your household ( have | own ) ?( any ) mosquito nets that can be used while sleeping
Variant do you ( have | own ) ?( any ) mosquito nets that can be used while sleeping
Variant do you use mosquito nets ( when you sleep | while you are sleeping )
Answers yes no

EndQuestion

# Questions for 'ShowNets'

Question
Group ShowNets
Translation/French Est-ce que je pourrais voir les moustiquaires ?
Translation/Spanish ¿Podría mostrarme los mosquiteros?
Variant ( can | could | may ) i ?( please ) see the ?( mosquito ) nets ?( in your household )
Variant ( can | could ) you ?( please ) show me the ?( mosquito ) nets
Answers yes no

EndQuestion

# Questions for 'BrandOfNet'

Question
Group BrandOfNet
Translation/French Quelle est la marque de votre moustiquaire ?.
Translation/Spanish ¿De qué marca es el mosquitero?
Variant what is the brand of ( this | the | your ) ?( mosquito ) net
Variant what brand is ( the | your | this ) ?( mosquito ) net
Answers brand_A brand_B brand_C dk_brand brand_D brand_E brand_F other

EndQuestion

# Questions for 'AnyoneSleptUnder'

Question
Group AnyoneSleptUnder
Translation/French Est-ce que quelqu'un a dormi sous cette moustiquaire la nuit dernière ?
Translation/Spanish ¿Durmió alguien bajo este mosquitero anoche?
Variant did anyone ( sleep under | use ) this ?( mosquito ) net ) ?( last night | yesterday )
Answers yes no dk/not_sure

EndQuestion

# Questions for 'WhoSleptUnder'
Question
Group WhoSleptUnder
Translation/French Qui a dormi sous cette mosquiquaire la nuit dernière ?
Translation/Spanish ¿Quiénes durmieron bajo este mosquitero anoche?
Variant who (slept under | used ) this ?(mosquito ) net last night
Answers name1 name2 name3
EndQuestion

# Questions for 'WallsSprayed'

Question
Group WallsSprayed
Translation/French Est-ce qu'au cours des 12 derniers mois, quelqu'un est venu dans votre logement pour pulvériser les murs intérieurs contre les moustiques ?
Translation/Spanish En algún momento durante los últimos 12 meses, ¿vino alguien a la casa para fumigar las paredes interiores con algún producto contra mosquitos?
Variant at any time in the past twelve months has anyone come into your dwelling to spray the interior walls against mosquitoes
Variant has anyone sprayed the ?(interior) walls ?(of your house) against mosquitoes within the past year
Variant in the past year has anyone sprayed the ?(interior) walls of your house against mosquitoes
Answers yes no
EndQuestion

# Questions for 'WhoSprayed'

Question
Group WhoSprayed
Translation/French Qui a pulvérisé le logement ?
Translation/Spanish ¿Quién fumigó la casa?
Variant who sprayed the (dwelling | walls of your house)
Answers government_worker/program private_company ngo other dk/not_sure
EndQuestion

ANSWERS

Answer
Content government_worker/program
PrintName Government worker / program
Code A
Translation/French Employé / Programme du gouvernement
Translation/Spanish Funcionario público / Programa del gobierno
EndAnswer

Answer
Content private_company
PrintName Private company
Code B
Translation/French Société privé
Translation/Spanish Empresa privada
EndAnswer

Answer
Content ngo
PrintName NGO
Code C
Translation/French ONG
<table>
<thead>
<tr>
<th>PrintName</th>
<th>Code</th>
<th>Translation/French</th>
<th>Translation/Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>DK / Not sure</td>
<td>98</td>
<td>Pas sûr</td>
<td>No está seguro</td>
</tr>
<tr>
<td>Other (specify)</td>
<td>96</td>
<td>Autre (préciser)</td>
<td>Otro (especifique)</td>
</tr>
<tr>
<td>Earth / Sand</td>
<td>11</td>
<td>Terre / Sable</td>
<td>Tierra / Arena</td>
</tr>
<tr>
<td>Dung</td>
<td>12</td>
<td>Bouse</td>
<td>Estiércol</td>
</tr>
<tr>
<td>Wood planks</td>
<td>21</td>
<td>Planches en bois</td>
<td>tablas de madera</td>
</tr>
<tr>
<td>Palm / Bamboo</td>
<td>22</td>
<td>Palmes/bambou</td>
<td>Palmera / Bambú</td>
</tr>
<tr>
<td>Parquet or polished wood</td>
<td>31</td>
<td>Parquet ou bois ciré</td>
<td>Parquet o madera lustrada</td>
</tr>
<tr>
<td>Vinyl or asphalt strips</td>
<td>32</td>
<td>Vinyle ou asphalte</td>
<td>Vinilo o asfalto</td>
</tr>
</tbody>
</table>
Answer
Content ceramic_tiles
PrintName Ceramic tiles
Code 33
Translation/French Carrelage
Translation/Spanish Cerámicos
EndAnswer

Answer
Content cement
PrintName Cement
Code 34
Translation/French Ciment
Translation/Spanish Cemento
EndAnswer

Answer
Content carpet
PrintName Carpet
Code 35
Translation/French Moquette
Translation/Spanish Alfombra
EndAnswer

Answer
Content in_the_house
PrintName In the house
Code NA
Translation/French Dans la maison
Translation/Spanish En la casa
EndAnswer

Answer
Content separate_room
PrintName In a separate room used as a kitchen
Code 1
Translation/French Dans une pièce séparée utilisée comme cuisine
Translation/Spanish En una habitación separada utilizada como cocina
EndAnswer

Answer
Content in_a_separate_building
PrintName In a separate building
Code 3
Translation/French Dans un bâtiment séparé
Translation/Spanish En una edificación separada
EndAnswer

Answer
Content outdoors
PrintName Outdoors
Code 4
Translation/French À l'extérieur
Translation/Spanish Al aire libre
EndAnswer

Answer
Content piped_into_dwelling
PrintName Piped into dwelling
Code 11
<table>
<thead>
<tr>
<th>Code</th>
<th>Content</th>
<th>PrintName</th>
<th>Translation/French</th>
<th>Translation/Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Piped to yard/plot</td>
<td>Piped into compound yard or plot</td>
<td>Dans le logement</td>
<td>Tubería a la vivienda</td>
</tr>
<tr>
<td>13</td>
<td>Piped to neighbour</td>
<td>Piped to neighbour</td>
<td>Dans concession, cour ou parcelle</td>
<td>Tubería al terreno, patio o lote</td>
</tr>
<tr>
<td>14</td>
<td>Public tap/standpipe</td>
<td>Public tap / standpipe</td>
<td>Robinet public / Borne fontaine</td>
<td>Llave/grifo público</td>
</tr>
<tr>
<td>21</td>
<td>Tube well or borehole</td>
<td>Tube well or borehole</td>
<td>Puits à pompe / Forage</td>
<td>Pozo con tubería</td>
</tr>
<tr>
<td>31</td>
<td>Protected well</td>
<td>Protected well</td>
<td>Puits protégé</td>
<td>Pozo recubierto</td>
</tr>
<tr>
<td>32</td>
<td>Unprotected well</td>
<td>Unprotected well</td>
<td>Puits non protégé</td>
<td>Pozo no recubierto</td>
</tr>
<tr>
<td>41</td>
<td>Protected spring</td>
<td>Protected spring</td>
<td>Source protégée</td>
<td>Manantial recubierto</td>
</tr>
</tbody>
</table>
Content unprotected_spring
PrintName Unprotected spring
Code 42
Translation/French Source non protégée
Translation/Spanish Manantial no recubierto
EndAnswer

Answer
Content rainwater_collection
PrintName Rainwater collection
Code 51
Translation/French Eau de pluie
Translation/Spanish Agua de lluvia
EndAnswer

Answer
Content tanker_truck
PrintName Tanker truck
Code 61
Translation/French Camion-citerne
Translation/Spanish Camión cisterna
EndAnswer

Answer
Content cart_with_small_tank
PrintName Cart with small tank / drum
Code 71
Translation/French Charrette avec petite citerne / tonneau
Translation/Spanish Carreta con tanque/tambor pequeño
EndAnswer

Answer
Content surface_water
PrintName Surface water (river, stream, dam, lake, pond, canal, irrigation channel)
Code 81
Translation/French Eau de surface (rivière, fleuve, barrage, lac, mare, canal, canal d'irrigation)
Translation/Spanish Agua de supreficie río, arroyo, represa, lago, estanque, canal, canal de irrigación)
EndAnswer

Answer
Content bottled_water
PrintName Bottled water
Code 91
Translation/French Eau en bouteille
Translation/Spanish Agua embotellada
EndAnswer

Answer
Content in_own_dwelling
PrintName In own dwelling
Code 1
Translation/French Dans logement
Translation/Spanish En el interior de la propia vivienda
EndAnswer

Answer
Content in_own_yard/plot
PrintName In own yard / plot
Code 2
Translation/French Dans cour / parcelle
Translation/Spanish En el propio patio / lote
EndAnswer

Answer
Content elsewhere
PrintName Elsewhere
Code 3
Translation/French Ailleurs
Translation/Spanish En otra parte
EndAnswer

Answer
Content flush/pour flush
PrintName Flush / Pour flush
Code FLU
Translation/French Chasse d'eau avec ou sans réservoir d'eau
Translation/Spanish Chorro / baldeo
EndAnswer

Answer
Content ventilated_improved_pit_latrine
PrintName Ventilated improved pit latrine
Code 21
Translation/French Latrines améliorées ventilées
Translation/Spanish Letrina de fosa mejorada con ventilación
EndAnswer

Answer
Content pit_latrine_with_slab
PrintName Pit latrine with slab
Code 22
Translation/French Latrines à fosses avec dalle
Translation/Spanish Letrina de fosa con losa
EndAnswer

Answer
Content pit_latrine_without_slab/open_pit
PrintName Pit latrine without slab / open pit
Code 23
Translation/French Latrines à fosses sans dalle/ Trou ouvert
Translation/Spanish Letrina de fosa sin losa / Foso abierto
EndAnswer

Answer
Content composting_toilet
PrintName Composting toilet
Code 31
Translation/French Toilettes à compostage
Translation/Spanish Inodoro de compostaje
EndAnswer

Answer
Content bucket
PrintName Bucket
Code 41
Translation/French Seaux
Translation/Spanish Balde
EndAnswer
Translation/Spanish No sabe marca
EndAnswer

Answer
Content brand_D
PrintName Brand D
Code 11
Translation/French Marque D
Translation/Spanish Marca D
EndAnswer

Answer
Content brand_E
PrintName Brand E
Code 11
Translation/French Marque E
Translation/Spanish Marca E
EndAnswer

Answer
Content brand_F
PrintName Brand F
Code 11
Translation/French Marque F
Translation/Spanish Marca F
EndAnswer

Answer
Content name1
PrintName Name 1
Code 1
Translation/French Nom 1
Translation/Spanish Nombre 1
EndAnswer

Answer
Content name2
PrintName Name 2
Code 2
Translation/French Nom 2
Translation/Spanish Nombre 2
EndAnswer

Answer
Content name3
PrintName Name 3
Code 3
Translation/French Nom 3
Translation/Spanish Nombre 3
EndAnswer

Answer
Content thanks
PrintName Thanks!
Code TK
Translation/French Merci !
Translation/Spanish ¡Gracias!
EndAnswer