Creating a Corpus of Spoken Modern Western Armenian

Marie Nilsenová
University of Tilburg
The Netherlands

Melanie Keledjian
INALCO, Paris
France

1 Goal of the project

The goal of our project is to collect data for a balanced multi-purpose corpus of spoken Modern Western Armenian, containing 27 different 4,000-word conversation fragments and 30 read scripted samples of sentences with phonetically varied sounds (300 words each), sampled with a non-probabilistic technique (a combination of "judgement" and "convenience" sampling, used, e.g., in the International Corpus of English). Since spontaneous conversations exhibit a high rate of internal variation, in order to collect representative samples, they will vary across the following variables: gender, age, level of education, dialect, social context and relationships (disparate/equal, intimate/non-intimate). The digitized recordings in the 'aiff' format (with an uncompressed sound) which will constitute the corpus will later be transcribed in UNICODE.

2 Why are corpora crucial for linguistic research?

A well-constructed corpus of a sufficient size is the main precondition for methodologically sound linguistic research to be carried out. Nowadays, language corpora are used as the basis for reference grammars, they provide natural non-contrived examples for text books and serve as an enhancement of foreign language teaching in "data-driven learning". In theoretical linguistics, they provide data to study language variation, specific linguistic constructions and language acquisition. While pure introspection is what drives research in this area of linguistics, it is a highly disputed method if the thus collected data serve to confirm or disprove a theory. As noted by [Labov, 1972]:199 and further discussed by [Schütze, 1996]:5, "linguists cannot continue to produce theory and data at the same time" because nothing can stop them from manipulating the introspection process to substantiate their own theories (knowingly or unknowingly). Luckily enough, thanks to the current technical developments, automated corpus studies have become highly reliable and less time-consuming than in the past.

While corpora of written texts are often easier to construct (and undoubtedly useful), spoken corpora reflect more truthfully the use of a language: only a small segment of the speakers of a language creates literary texts and news reports that written text corpora are based on and spontaneous dialogues are much more frequent than written production. (It has been noted that a strictly 'proportional' corpus would have to contain roughly 90% conversations.) While *read speech* corpora are usually of a higher quality than *free conversational speech* corpora, they may lack in naturalness and representativeness. The use of free conversational speech corpora – if of acceptable quality – appears to be preferable in that it ensures ecological validity of the studied examples (i.e., that they closely approximate real life use).

3 Why Modern Western Armenian, why in Syria?

There is currently no existing spoken corpus for Modern Western Armenian that could be used in the sense described above. Given the complex socio-linguistic situation in which the language has evolved, empirical data regarding its current state are of high interest to Indo-European language studies. At the same time, the corpus will contribute real-life examples in the context of second language learning abroad, driven by the desire to preserve a cultural heritage whose heirs have mostly been living dispersed around the world.

Out of the few communities around the globe where it is spoken, the Syrian variant of Modern Western Armenian has been selected as the representative one for the corpus. Unlike in Europe and in America, the development of of Modern Western Armenian in the Middle East has been continuous (no generations of speakers had to learn the language later in life from textbooks, as it is currently the case in the U.S.).

4 An intended use of the corpus: intonation research

Most corpus-based theoretical research focuses on lexical and morphosyntactic phenomena, but a spoken corpus is especially suitable for the research of purely sound-based – yet highly meaningful – linguistic features, such as intonation. Intonation is normally defined as a combination of melody and intensity, though in practice the term is often used to refer just to the fundamental frequency (or *pitch*, its psychophysical correlate) of the signal. How do speakers produce fundamental frequency/pitch? Basically, when we talk, the airflow pressed out from our lungs travels through the trachea into the nasal and oral cavity and out into the open. The air passing through the (open) vocal chords causes them to vibrate. The vibration can be slower or faster, depending on the 'size' of the folds (their mass) and the tension in the muscles that control their use; simplifying somewhat, the higher the tension and the smaller the size, the higher the frequency with which air pulses will be released into the vocal tract, and the higher the pitch. Given that smaller people, women and children tend to have smaller and thinner vocal folds and shorter vocal tracts, the pitch they produce is generally higher (between 180 and 400 Hz for women)[1] than that of larger people

[1]Most frequently, pitch is expressed in terms of Hz (number of vocal cord opening and closing events per second) or in semitones (a logarithmic unit which preserves the distance magnitude between frequencies independently of

and/or males, with some dependence on age (the typical male voice is somewhere between 60 and 240 Hz). This fact, together with our ability to modulate pitch, lies behind the so-called Frequency Code discussed in more detail below. The intonational properties of an utterance can be visualized as in figure 1, where the thick line indicates the rising and falling pitch contour, drawn into the intensity diagram (with intensity represented by the thin line) of the signal.

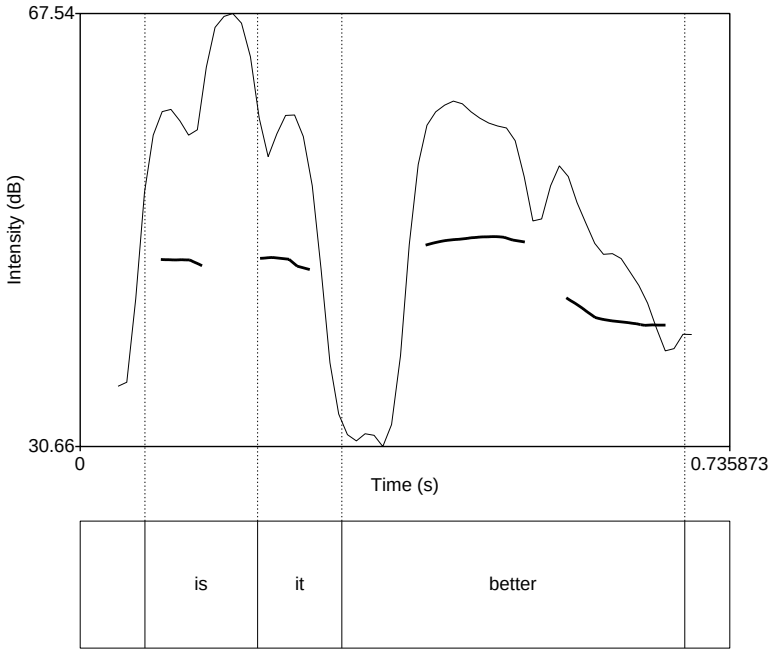

Figure 1: Visualizing the intonational properties of an utterance - pitch (thick line) drawn into an intensity diagram (thin line).

Fundamental frequency plays a very important role in communication and the human ear is particularly well adapted to distinguish frequencies in the 100-500 Hz range (for longer and purer tones, down to the difference of 2 Hz). It seems that most (if not all) languages exploit pitch changes to signal paralinguistic information, where the term 'paralinguistic' refers to emotions and attitudes such as sadness, joy, surprise, irritation, anger, sarcasm, fear, and so on. It has sometimes been assumed that conveying the paralinguistic meaning is the main role of intonation. There is enough linguistic evidence, though, showing that the use of intonation is at least partly grammaticalized and the association to emotions exists in parallel to its linguistic function, which in some cases is directly derived from the paralinguistic use. From a linguistic point of view, languages employ pitch in different ways (although universal tendencies can be found). Some Indo-European languages use so-called lexical accent, i.e., pitch on the level of words or phrases, to signal lexical or part-of-speech differences (e.g., Norwegian). Other languages – like Mainstream American English, Standard French or Standard Dutch – apart from word stress use pitch mainly to signal properties of larger discourse segments (utterances, paragraphs). Compared to lexical accents, the use of pitch on this level appears to be more difficult to capture in linguistic terms because it may associate with segments of various types and lengths.

speaker's range, viz. ['t Hart et al., 1990]), though other scales can also be used.

It has been argued that languages share some universal tendencies in the use of intonation and intonational universals can provide for interesting hypotheses.[2] An example of what has long been assumed to be an intonational universal is the supposedly wide-spread tendency for questions to be signaled with rising intonation (70 % of a sample of 250 languages, viz. [Bolinger, 1978]). One problem with the generalization is that the claim concerns only *yes/no*-questions, and even for this class it is not based on corpus or experimental results. In languages for which corpus studies have been done (such as English, viz. [Fries, 1964] and [Hirschberg, 2000], or French, viz. [Fónagy and Bérard, 1973]), the original observation had to be modified because in some corpora, less than 40 % of spontaneous polar questions were, in fact, rising. Further crosslinguistic evidence was recently offered by [Rialland, 2004], who observes that in a database of 80 African languages, questions without any pitch raising, high tones or rising intonation were quite common (e.g., in the Gur languages, as well as among the Kwa, Kru and Mande languages). Also, even in languages with rising questions, the contour may not be considered the canonical polar question contour.

Apart from linguistic data, universal claims regarding intonation use have been based on physiological facts about intonation production. For example, as reported in [Chen, 2004]:2, it has been observed that the degree of perceived emotion (bored-interested, calm-excited, afraid-bold, etc.) is positively correlated with intonation across unrelated languages, and suggests that this is due to a universal factor behind the physiology of emotion and voice production (emotion leads to an increase in muscular tension and activity and hence to higher pitch). [Ohala, 1983] and [Ohala, 1984], on the other hand, proposes an ethological basis for the use of pitch which makes use of the observation that in the animal kingdom, deep sounds are interpreted as threatening, while high pitched sounds are interpreted as weak and submissive.

Building partly on the work of Ohala, Gussenhoven ([Gussenhoven, 2002], [Gussenhoven, 2004]) identifies three biological codes related to the use of pitch for signalling information (both paralinguistic and linguistic). The three codes, called the *Frequency Code*, the *Effort Code* and the *Production Code*, are based on the following observations (as formulated in [Gussenhoven, 2002]):

> **The Frequency Code.** Smaller larynxes contain lighter and smaller vocal cords, with which faster vibration rates are achieved for a given amount of energy. The correlation between larynx size and rate of vocal cord vibration is exploited for the expression of power relations across species: lower pitch is associated with larger body size and hence with social dominance ([Ohala, 1983], [Ohala, 1984]).
>
> **The Production Code.** The generation of energy is tied to the exhalation phase of the breathing process, and hence becomes available in phases. As a consequence, high pitch is associated with the beginnings of utterances, while low pitch is associated with the ends.
>
> **The Effort Code.** The amount of energy expended on speech production can be varied: putting in more effort will not just lead to more precise articulatory movements, but also to more canonical and more numerous pitch movements. Lavishing more care on the production process means less slurring together of these movements, causing

[2] The field of intonational typology is recently rapidly developing ([Vaissière, 1995], [Hirst and DiCristo, 1998]), however, no typologically representative sample of languages from different languages groups and families has so far been collected.

them to be carried out with less undershooting of targets.

Gussenhoven's hypothesis is interesting because, among other things, it relates phonetic and phonological aspects of pitch use and makes specific predictions regarding their linguistic interpretation. With respect to the relation between phonetics and phonology, the biological codes are assumed to be phonetically implemented in the paralinguistic production and interpretation of pitch in all languages[3] and to have an effect on the phonology in most. While linguistic intonational meaning may diverge from the canonical interpretation provided by the codes as a result of language change, overall, languages are expected to exhibit the tendency to use grammatical categories directly derived from the paralinguistic meaning. Moreover, even in those languages where the linguistic implementation of intonational categories goes against the biological codes, speakers will still exploit the original phonetic effect to convey their attitudes and emotions.

For example, the Frequency Code is based on Ohala's claim that there is a correlation between larynx size, fundamental frequency and body size, which is used for expressing power relations. In Gussenhoven's interpretation, the code's paralinguistic interpretations are for high pitch 'submissiveness', 'femininity', 'politeness','vulnerability' and 'friendliness', whereas for low pitch it is 'dominance', 'masculinity', 'confidence' and 'aggression'. On the level of linguistic information, higher pitch is supposed to convey 'uncertainty' (as opposed to 'certainty') and hence 'questioning' (as opposed to 'asserting'), based on the idea that "when asking questions, one is dependent on the other's good will for the information requested" ([Chen, 2004]:33).

However, for an analysis of intonational meaning to be possible, information about acoustic properties of the continuous speech signal has to be reduced into linguistically relevant units, i.e., it has to be *symbolically transcribed*. Despite repeated attempts, there is so far no fully automated prosodic transcription available; human intervention is required on all levels of the prosodic analysis which can make it costly and (at least partly) subjective. There are many transcription systems around which differ both in their basic assumptions regarding intonational morphology, as well as in its symbolization. Below, some examples of intonation transcription systems are given. In the notation of [Cruttenden, 1997] in figure 2, the top and bottom lines represent speaker's minimum and maximum pitch and each dot stands for a syllable (larger dots indicate an accented syllable). In the so-called 'close copy stylization' transcription, on the other hand, the pitch contour is reproduced as a series of dots. Bolinger's technique (figure 3), which has been referred to as 'scrolling typewriter', uses the transcription of the utterance to mirror the contour. Also, 'down arrows' and 'up arrows' are frequently employed to indicate rising and falling movements. The same - and more - can be done with letters, as in the INTSINT alphabet of [Hirst and DiCristo, 1998], where letters are used as symbols for tonal changes together with local minima and maxima (figure 4).

INTSINT (INternational Transcription System for INTonation) is a language-independent intonation transcription system developed in Aix-en-Provence (viz. [Hirst and DiCristo, 1998]). It is frequently used to transcribe French intonation but it has also been employed for other languages (viz. Hirst & Di Cristo's volume on intonational typology [Hirst and DiCristo, 1998]) and it is our intention to use it also for the Modern Western Armenian corpus under construction. INTSINT labels target points (tones) determined by the accompanying MOMEL (MOdélisation de MELodie)

[3][Chen, 2004] examined the paralinguistic interpretation of the codes in a series of perceptual experiments and found varying tendencies in judgements of speakers of different languages. I will leave this point aside here, though, since I am mainly concerned with the linguistic meaning.

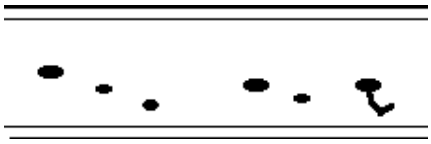

Figure 2: The ‘interlinear tonetic transcription’ (a.k.a. ‘dots with tails’).

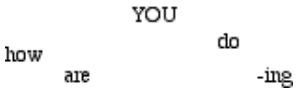

Figure 3: Bolinger’s ‘scrolling typewriter’ technique.

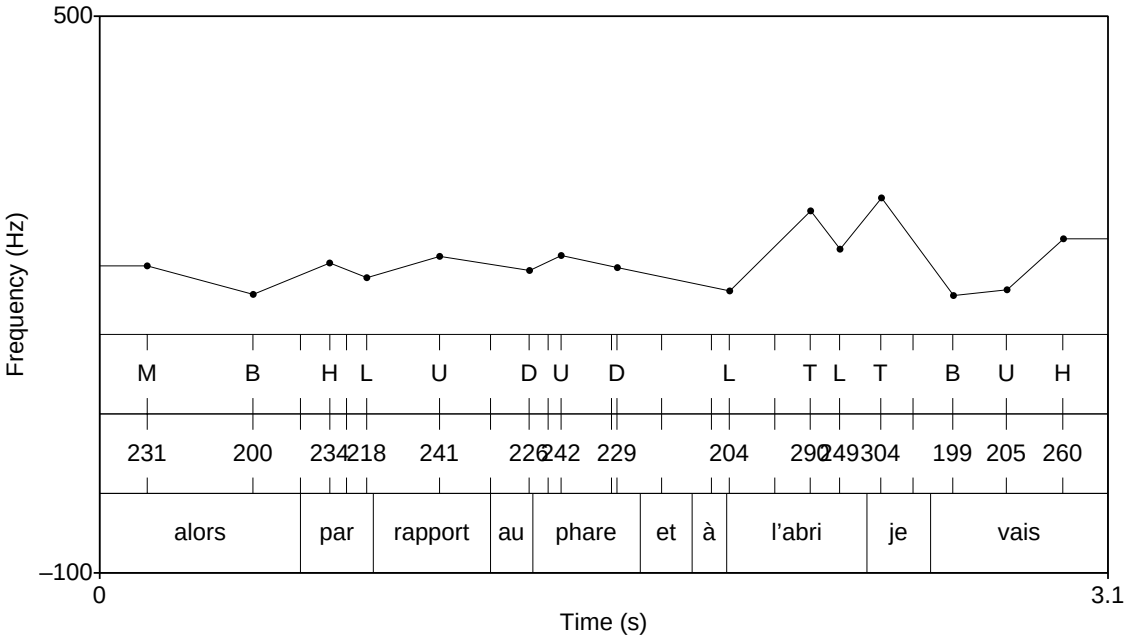

Figure 4: A French utterance with a MOMEL curve and INTSINT annotation.

algorithm. It annotates both absolute and relative prosodic events (making use of the MOMEL modelled pitch curve), with the following set of symbols:

- **T** – Top; **M** – Mid; **B** – Bottom; **H** – Higher; **S** – Same; **L** – Lower; **U** – Upstep; **D** – Downstep

The coding can be done automatically in the freely available PRAAT program (www.praat.org). An example of a part of a French utterance annotated with the INTSINT alphabet and the stylized MOMEL curve is given in the figure 4, with the INTSINT annotation in the second grid from above. The resulting annotation can be evaluated and compared to data found in French and other languages.

5 Summary

The primary goal of our project is to create a reasonably balanced corpus of conversational speech produced by native speakers of Modern Western Armenian, which would serve a variety of theoretical and applied linguistic purposes. The secondary goal is to create a source of empirical data for typological intonation research, in order to test claims regarding the universality of intonation use and meaning on a language that so far remains largely unexplored from this perspective.

References

[Bolinger, 1978] Bolinger, D. (1978). Intonation across languages. In Greenberg, J., editor, *Universals of Human Language*, volume II: Phonology, pages 471–524. Stanford University Press, Palo Alto, CA.

[Chen, 2004] Chen, A. (2004). *Universal and Language-specific Perception of Paralinguistic Intonational Meaning*. PhD thesis, LOT dissertation series 102.

[Cruttenden, 1997] Cruttenden, A. (1997). *Intonation*. New York: Cambridge University Press, 2nd edition edition.

[Fónagy and Bérard, 1973] Fónagy, I. and Bérard, E. (1973). Questions totales simples et implicatives en Français Parisien. In Grundstrom, A. and Léon, P., editors, *Interrogation et Intonation*, number 8, pages 53–98. Didier, Paris.

[Fries, 1964] Fries, C. C. (1964). On the intonation of yes-no questions in English. In Abercombie, D., Fry, D., McCarthy, P., Scott, N., and J.L.M. Trim, L., editors, *In Honour of Daniel Jones*. Longman.

[Gussenhoven, 2002] Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In Bel, B. and Marlien, I., editors, *Proceedings of Speech and Prosody '02*, Aix-en-Provence.

[Gussenhoven, 2004] Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.

[Hirschberg, 2000] Hirschberg, J. (2000). A corpus-based approach to the study of speaking style. In Horne, M., editor, *Prosody: Theory and Experiment*, pages 335–350. Kluwer Academic Press.

[Hirst and DiCristo, 1998] Hirst, D. and DiCristo, A., editors (1998). *Intonation systems: a survey of twenty languages*. Cambridge University Press.

[Labov, 1972] Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

[Ohala, 1983] Ohala, J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, (40):1–18.

[Ohala, 1984] Ohala, J. (1984). An ethological perspective on common cross-language utilization of f0 in voice. *Phonetica*, 41:1–16.

[Rialland, 2004] Rialland, A. (2004). A typology of question prosody in african languages. presented at the International Conference on Tone and Intonation.

[Schütze, 1996] Schütze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgements and Linguistic Methodology*. University of Chicago Press.

['t Hart et al., 1990] 't Hart, J., Collier, R., and Cohen, A. (1990). *A Perceptual Study of Intonation - An experimental-phonetic approach to speech melody*. Cambridge University Press, Cambridge.

[Vaissière, 1995] Vaissière, J. (1995). Phonetic explanations for cross-linguistic prosodic similarities. *Phonetica*, (52):123–130.