EXPLORING PRONUNCIATION VARIANTS FOR ROMANIAN SPEECH-TO-TEXT TRANSCRIPTION

I. Vasilescu[1], B. Vieru[2], L. Lamel[1]

[1]LIMSI-CNRS - B.P. 133 91403 Orsay France,
[2]Vocapia Research - 28 rue Jean Rostand, 91400 Orsay France
ioana@limsi.fr, bianca@vocapia.com, lamel@limsi.fr

ABSTRACT

Speech processing tools were applied to investigate morpho-phonetic trends in contemporary spoken Romanian, with the objective of improving the pronunciation dictionary and more generally, the acoustic models of a speech recognition system. As no manually transcribed audio data were available for training, language models were estimated on a large text corpus and used to provide indirect supervision to train acoustic models in a semi-supervised manner. Automatic transcription errors were analyzed in order to gain insights into language specific features for both improving the current performance of the system and to explore linguistic issues. Two aspects of the Romanian morpho-phonology were investigated based on this analysis: the deletion of the masculine definite article *-l* and the secondary palatalization of plural nouns and adjectives and of 2nd person indicative of verbs.

Index Terms— ASR, Romanian, speech transcription errors, pronunciation variants, definite article, palatalization.

1. INTRODUCTION

With increasing globalization, there is a growing need to bridge the numerical gap between technologically privileged countries with the developing world. In this framework, developing language technologies for under-resourced languages is challenging. For the particular case of the enlarged political Europe, there is a pressing need to extend language technologies to less studied European languages. Romanian figures among these European languages, as Romania joined the European Union in 2007.

Automatic speech recognition (ASR) systems have mainly been developed for the world's dominant languages (e.g., English, Arabic, Chinese, French) for which large transcribed speech corpora, as well as plethora of written texts are available in electronic form. Traditionally, such systems are trained on large amounts of carefully transcribed speech data and very large quantities of written texts. However obtaining large volumes of transcribed audio data remains quite costly and requires a substantial time investment and supervision.

Studies addressing large vocabulary continuous speech recognition for the Romanian are lacking, however some attempts to build STT systems on small corpora have been reported recently [1], [2]. As part of the speech technology development in the Quaero program [1], a Romanian ASR system targeting broadcast and web audio was built. The acoustic models were developed in an unsupervised manner as no detailed annotations were available for the audio training data downloaded from a variety of websites similar to the method employed in [3].

The automatic transcripts have also been used for further linguistic studies to gather information about specific trends of spoken Romanian. As previous studies have pointed out, looking into automatic speech transcription errors may provide precious insight about contextual variation in conversational speech and more generally about potential systematic mutations [4, 5]. In this work, automatic alignment of speech data with dictionaries containing specific pronunciation variants is used to investigate two instances of variation illustrated by the automatic transcription errors in Romanian: the deletion of the masculine definite article *-l* and the secondary palatalization of plural nouns and adjectives, and the 2nd person indicative form of verbs.

To the best of our knowledge, published linguistic studies on spoken Romanian have been carried out using small controlled corpora of prepared speech (e.g. sentences recorded in laboratory conditions). For instance, a recent study on the production and perception of word-final secondary palatalization pointed out differences in the acoustic realization of the palatalization according to the manner and place of articulation of the consonant [6]. In contrast, the study described in this paper is conducted on a mix of semi-prepared and conversational speech resulting in 3.5 hours of broadcast data.

The next section gives a brief description of the Romanian language. Section 3 is dedicated to a description of the corpora used to train and develop the Romanian ASR system, followed by an overview of the ASR system and models and recognition results on Quaero data. Section 4 presents an ASR error analysis carried with the development data. Based

---

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

on the observed errors, two case studies of contextual variation are investigated and described in Section 5.

2. BACKGROUND: ROMANIAN LANGUAGE

This section focuses on a general description of the Romanian language and history and highlights some of the potential challenges for speech recognition.

2.1. Description of the language

Spoken by over 29 million speakers around the world, Romanian is mother tongue of 25 million of them and is the official language of two countries: Romania and Republic of Moldavia [7]. Romanian is a Latin language, from the Oriental branch, developed at a geographical distance from the other Romance languages and surrounded by Slavic languages. For this reason, Romanian shows some peculiarities compared to the other Romance languages. Romanian is based on late Vulgar Latin, having been among the last territories conquered by the Roman Empire. Due to its geographical isolation, elements of the Vulgar Latin have been accurately preserved: for instance, modern Romanian has preserved the Latin declension of nouns and verbs, a feature lost in the other Romance languages. The majority of the basic vocabulary is of Latin origin (about 60%), however Romanian has borrowed numerous features from the Slavic languages with which it has been and still is in contact. Such features may be encountered at various linguistic levels: phonetics, lexical, morphology and syntax, however, most aspects of Romanian morphology remain close to Latin. The Slavic influence was reinforced by the long-term use of the Cyrillic alphabet (introduced in Romania with the oriental Christian religion and adapted to written Romanian). After the 18th century, Romanians, proud of their Latin origins, borrowed many words from Romance languages such as French and Italian. As a result, in the history of the Romanian language a "re-Latinization" of the language occurred [8], [9]. Finally, political, economic and social aspects in the history of Romania explain other Eastern European influences: Turkish, Greek, German, Hungarian etc. Today, the English language is having a strong influence on the Romanian language at the lexical level. To sum up, Romanian may be described as a Latin language (phonetic, morpho-syntactic and lexical levels), with strong Slavic influences (phonetic and lexical levels) and also with increasing elements from contemporary Romance and English languages (lexical level).

2.2. Voice technologies for Romanian

Romanian shows specific features which can be a challenge for ASR systems and language technology in general. As an example, Romanian is a highly inflected language, with various specific patterns: it is a pro-drop language (as most of the Romance languages allowing the deletion of the subject), it allows clitic doubling, negative concord and double negation [8].

One of the main problems in automatically modeling the language is the richness of the inflection inherited from Latin. For nouns, pronouns and adjectives there are 5 cases. However some reductions occur, for nouns two oppositions being functional, Nominative/Accusative forms vs Dative/Genitive, the Vocative being possible but sharing often the same form as Nominative. Pronouns can have stressed and unstressed forms, while nouns and adjectives can be definite and indefinite. There are 3 genders (masculine, feminine and neutral, the latter sharing some forms with masculine and feminine and being partially predictable [10]). For verbs there are two numbers, each with three persons, and five synthetic tenses, plus infinitive, gerund and participle forms. In average, a noun can have 5 forms, a personal pronoun and an adjective about 6 forms, while a verb may have more than 30 forms. However, in spontaneous speech some opposition may be neutralized, in particular when the affixes are word final, thus less carefully articulated, as the current work will demonstrate. Besides morphological suffixes and endings, phonetic alternations inside the root are also possible with inflected words [7].

Romanian belongs to the new EU languages poorly represented in the speech technology world whereas the presence of their speakers across enlarged Europe represents a real challenge for such technologies. For instance, according to [7] automatic speech recognition is one of the less represented voice-driven technology dedicated to Romanian language.

3. ASR SYSTEM FOR ROMANIAN

The section describes the data used for building and testing the Romanian ASR system. Normalization issues specific to Romanian language are discussed as well.

3.1. Data description

Within the Quaero program a corpus of 3.5 hours of speech with detailed manual transcriptions was distributed as development data. This corpus was used in the experiments reported below. The corpus annotation team selected the data from several broadcast/podcast sources so as to cover the various styles of speech found in such shows, from read speech to more spontaneous interactions, that is interviews. The speech segments were taken from 88 different speakers, with attention to have a balance for the amount of data per speaker and in terms of male and female (38%) speakers. As this data was to be used for initial work on Romanian, attention has been paid to avoid audio files that were highly interactive (too many segments with overlapping speech), or had strong foreign or regional accents, noisy background etc, as such data

Source	#Audio	#speakers
trn1	52h	1.5k
trn2	350h	23k
dev	3.5h	88

Table 1. Duration and number of speakers for the transcription used for training (trn1 and trn2) and for development and linguistic analysis (dev). For the training data the speaker counts result from the partitioning stage.

types remain challenging for state-of-the-art ASR systems for well-resourced languages. The data are representative of the 'standard' version of the language which is mainly spoken in southern Romania. For the linguistic investigations reported in Section 5 the data come from BN sources, that is two radio stations (RFI Journal, RRA - Radio Romania Actualitati), a news agency (Euranet) and a tv station (Antena 3). The latter source consists of debates whereas the first three correspond to news, interviews and short debates. The number of speakers varies from 3 (Euranet) to 24 (Antena 3) per audio source.

An additional 400+ hours of audio data from two broadcast sources (trn1 and trn2) were used for acoustic model training. The data are summarized in Table 1. For the dev data the speaker count was obtained from the manual transcriptions, however for the training audio these counts are only approximate, and certainly superior to the real number of distinct speakers, as they are obtained by the audio partitioner [11] and were not clustered across audio files.

3.2. Text Normalization

For the text data, each paragraph was first tested to determine if it is really in the Romanian language or not, and then the text was cleaned removing html tags and special characters. Numbers have been transformed in letters to be closer to the spoken form.

Special care was paid to diacritics. Indeed diacritics change words meanings and their pronunciations. For a native speaker, the text remains comprehensible if the diacritics are missing, but the text is not orthographically correct. Almost half of the textual data that were downloaded do not have diacritics. Many words differ only with respect to diacritics and these induce pronunciation changes. For example the endings /a/-/ă/ mark the difference between words with a +/- definite article. Two possibilities were considered to address this problem:

1. consider the two variants in the recognition dictionary as pronunciation variants,

2. automatically add diacritics to the text sources that did not have them.

As very few manual transcriptions were available we considered the first option to be less appropriate for this work.

Consequently, diacritics were automatically added to the texts that did not include them.

A word distance was used in order to select equivalent words creating a word correspondence list of 19k words which was manually filtered. The word distance is calculated based on the Levenshtein distance which is a string metric for measuring the difference between two sequences. The distance between two words is the minimum number of single-character edits (i.e. insertions, deletions or substitutions) required to change one word into the other, where the insertions and deletions are more costly than the substitutions. For example, the distance between *cercetări cercetari "research"* is 1 and the distance between *cercetători "researchers"* and *cercetari "research"* is 5. This equivalence list and the text sources containing correct diacritics were used to to build a language model, which was then used to add diacritics to the texts that were lacking them. For instance, words such as *cercetări cercetari "research"* and *cetăţean cetatean cetăţean cetaţean (citizen"* are considered equivalent and are mapped to the first variant.

3.3. Phone set

The Romanian language is written using the Latin script, making use of 31 letters with five additional letters â, î, ă, ş, ţ. The letters â and î play the same phonetic role, both being in use for historical reasons. Except minor exceptions (*x* is pronounced as /ks/, *ce* and *ci* as /tʃ/...), Romanian is a phonemic language. Although in the spoken language palatalization and poorly articulated affixes can be observed, only simple rules corresponding to "standard" language were applied for the grapheme-to-phoneme conversion.

The phone set used for the Romanian system contains 33 symbols: 20 consonants, 3 semi-vowels, 7 vowels and 3 special symbols (see Table 2). The correspondence between letters and phones is almost one-to-one. About twenty rules were used to transform letters into phones. There are on average 1.04 pronunciations/word. The pronunciations of some frequent words of foreign origin were manually corrected.

3.4. ASR system and results

The recognizer uses the same basic statistical modeling techniques and decoding strategy as in the LIMSI English BN system [11]. Language models (2, 3, 4-grams) were built using 79 millions of words from the normalized texts. The texts were accumulated from 6 different websites, with varying quantities (between 100k words and 51M words). About 234K words correspond to texts that were considered as approximate transcriptions of audio data. A 130k word list was chosen by selecting the most likely words of the 1-gram model interpolated so as to minimize the perplexity on the development data. The out-of-vocabulary (OOV) rate measured on the development data is of 1.8%. A language model

IPA	Ex. Romanian	IPA	Ex. Romanian
p	**p**as	b	**b**an
t	**t**are	d	**d**acă
k	**c**al	g	**g**ol
m	**m**ic	n	**n**or
f	**f**oc	v	**v**al
s	**s**are	z	**z**id
h	**h**orn	t͡s	**ț**ara
r	**r**epede	l	**l**ung
ʃ	**ș**arpe	ʒ	**j**ar
tʃ	**c**er	dʒ	**g**er
a	**a**pa	e	**e**rou
i	**i**nsula	o	**o**ra
u	**u**da	ə	ud**ă**
ɨ	**î**nspre		
o̯a	f**oa**rte	j	**ia**pa
e̯a	m**ea**		
-	silence	-	breath
-	filler		

Table 2. Phones used in the Romanian ASR system

was constructed on each source with the Kneser-Ney discounting and then the individual models were interpolated so as to minimize the perplexity on the development data. The 4-gram model has a perplexity of 158 on the development data set.

Only 3.5 hours of manually transcribed data was at our disposal for developing the Romanian STT system. It was decided to reserve this data for development purposes and to develop the system in a semi-supervised manner as reported in [12, 3]. The process was started via language transfer, using acoustic seed models taken from other languages. Phones from French, Italian, Spanish and Polish were mapped to the Romanian phones, and phone models were extracted from corresponding acoustic model sets. About 15 minutes of audio was decoded using these seed models and a supposedly biased language model.

Data was added progressively in batches, with 10 iterations of decoding the increasing larger batches and estimating larger models to use in the next decoding round. The acoustic models of the last iteration were trained on about 400 hours of speech.

Gender-dependent (GD) acoustic models were trained for MLP+PLP+F0 features [13]. No special MLP parameters were trained for the Romanian language: the MLP parameters trained for Italian were used [14]. All acoustic models are tied-state, left-to-right context-dependent, HMMs with Gaussian mixtures. The triphone-based context-dependent phone models are word-independent, but word position-dependent. The tied states are obtained by means of a decision tree where the 57 questions concern the phone position, the distinctive

features (vowel, consonant, nasal, stop, fricative, rounded, front, low, ... and identities) of the phone and the neighboring phones. Silence is modeled by a single state with 1024 Gaussians. The final model set covers about 15k phone contexts, with 11k tied states and 32 Gaussians per state.

The LIMSI/VOCAPIA Romanian ASR system obtained a word error rate (WER) of 17.1% on the 3.5 hour Quaero development corpus, and 19.9% on the Quaero 2012 evaluation data (official result). The WER range across the files is from 8.3% to 23.5%, with files containing more spontaneous speech having, as expected, a higher WER then those containing prepared speech.

4. ASR ERROR ANALYSIS

This section provides a brief description of the ASR errors as cues to language-specific ambiguities. Previous studies underlined that investigating automatic speech transcription errors may provide precious insights about the potential ambiguities of a language [4]. Transcription errors underline speech regions which are problematic for the ASR system [4]. These speech regions may correspond either to intrinsic ambiguities of the language or to some type of intra- and/or inter-speaker variation not properly accounted for in the system's speech model [15]. From a linguistic standpoint, the errors may be indicators of contexts and acoustic manifestations of variation. More largely, they may help in assessing if the observed variation is contextual and rather unpredictable or tends to generalize to systematic mutations [5]. From the ASR lexical modeling perspective they may point to appropriate pronunciation variants should be included in the dictionary.

Frequent confusions concern Romanian conjugation and declension peculiarities and in particular word final affixes. Such affixes, less carefully articulated and often subject to confusions. Among the pre-word elements, verbal subjunctive (*să* [sə]) and reflexive (*s(e)* [se]) marks behave as auxiliary elements: short and acoustically poor they may be easily deleted in continuous speech.

Verbal conjugation:

- (systematic) deletion of the mark of subjunctive *să*. The REF indicate the correct form and the HYP shows the system transcription.
  REF   să scoată to extract
  HYP   * scoată to extract

- (systematic) deletion of the reflexive mark *s(e)*
  REF   s-au aplecat they leaned over
  HYP   au aplecat leaned over

- auxiliary verb substitutions and deletions
  REF   a plecat he has left
  HYP   au plecat they have left

REF   a expulzat _{he ejected}

HYP   * expulzat _{ejected}

- other verbal tense/person confusions
  REF   au fost mai scurte _{they were shorter}
  HYP   a fost mai scurte _{he was shorter}
  REF   survola _{he was flying}
  HYP   survolau _{they were flying}

Nominal and adjectival declension:

- singular/plural substitutions:
  REF   reforme _{reforms}
  HYP   reformă _{reform}
  REF   uninominali _{reforms}
  HYP   uninominal _{for a single member}
  REF   acorduri _{accords}
  HYP   acordul _{accord}
  REF   sociali _{socials}
  HYP   social _{social}

- undefined and defined article deletion or substitution:
  REF   lista _{the list}
  HYP   listă _{a list}
  REF   soluţii _{solutions}
  HYP   soluţia _{the solution}
  REF   pentru o ţară _{for a country}
  HYP   pentru * ţară _{for country}

- case confusion e.g. Genitive/Dative with Nominative/Accusative:
  REF   monedei _{to the coin}
  HYP   monede _{the coin}
  REF   primului _{to the first}
  HYP   primul _{the first}

5. PATTERNS IN ROMANIAN INFLECTION & PRONUNCIATION VARIANTS

Automatic speech recognition has fostered the development of very large-scale spoken data with corresponding orthographic transcriptions [16]. The acoustic models can be used to generate segmentations of the speech into words and subword units, such as phones. Depending on the pronunciation dictionary's options and the acoustic model's accuracy, the resulting phone stream provides a more or less accurate phonetic or phonemic labeling. Beyond enabling the development of spoken language technologies, such data are also valuable for linguistic studies: the ASR systems may be used as a tool to highlight linguistic variation and to determine whether an observed phenomenon occurs randomly or follows regular patterns.

Linguistic investigation is based on a three step methodology:

- (1) The expected variation is first noticed as ASR transcription error;

- (2) The hypothesis of a variable realization is tested by authorizing pronunciation variants according to linguistic hypotheses for each case figure.

- (3) The results are (manually) verified : selected variants are analyzed in terms of (i) the acoustic realization of the expected phenomenon and (ii) the system's performance as an indicator of the effective realization of the hypothesized variation.

There is a lack of recent linguistic descriptions for the spoken Romanian language as most studies have been based on the written form. Consequently intra- and inter-speaker variation in the spontaneous speech, and more generally speaking, the phonetic variation in the contemporaneous Romanian has not been (well) studied. From the linguistic point of view, such studies may increase the knowledge of the phenomena which contribute to the evolution of a language. From the ASR point of view, accounting for variation may contribute to better modeling of pronunciation variants. Romanian is an interesting "use case" as the language shares both Romance and Slavic patterns. Hypothesis about intra- and inter-language pronunciation variants can be studied, in particular the influence of Russian, a Slavic language which historically influenced the Romanian phonetic system.

Two phenomena, both related to the phono-morphology of the language, are problematic for ASR systems, but for different reasons:

- the masculine definite article -l has a free variable realization as (*sistemu* or *sistemul* the system), and

- the final consonant palatalization marking inflections is very subtle and often times barely audible, possibly due to devoicing of the palatal articulation.
  plop – plopi [plop] – [plop^j] poplar – poplars;
  sap – sapi [sap] – [sap^j] I dig – you dig;
  ban – bani [ban] – [ban^j] money – monies

Both are word-final phenomena are susceptible to deletion in continuous speech. Figure 1 shows two spectrograms illustrating the absence (left) and presence (right) of the final l in the word *sistemul*. Figure 2 shows spectrograms of three renditions of the word *bani*, with palatalization in the rightmost example.
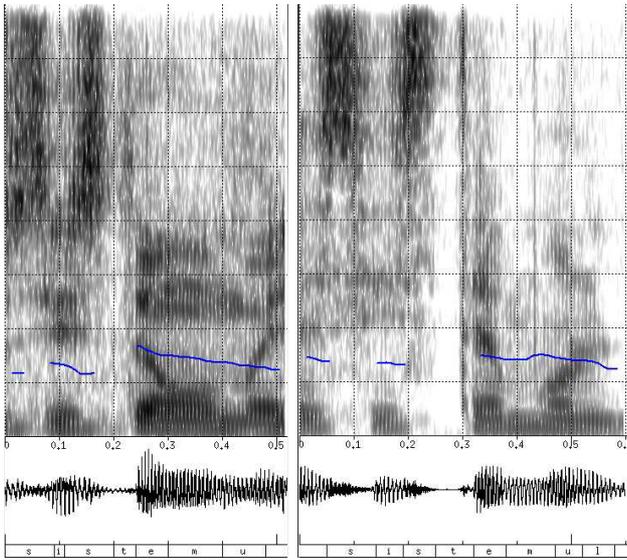
Fig. 1. Example spectorgrams of pronunciation variants for the word *sistemul* the system as selected by the ASR system: *sistemu* (left) vs *sistemul* (right). For each word, the horizontal axis shows the time aligned phonemes and the ticks on the vertical frequency axis are 1kHz (range 0-8kHz).
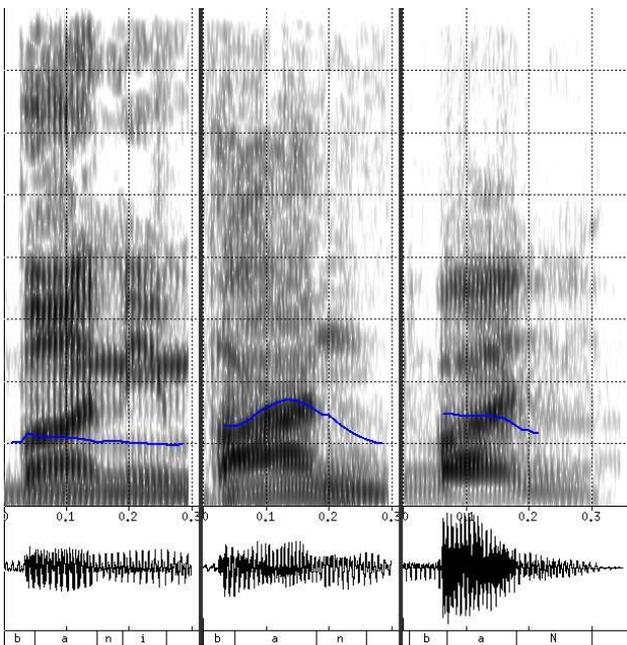
Fig. 2. Example spectrograms of pronunciation variants for the word *bani* monies as selected by the ASR system: The rightmost example is considered palatalized (N) vs the nasal (n) in the other two examples. For each word, the horizontal axis shows the time aligned phonemes and the ticks on the vertical frequency axis are 1kHz (range 0-8kHz).

5.1. Definite articles

The definite articles are attached to the end of the noun as enclitics:

*pom – pom***ul** [pom] – [pomul] tree - the tree;

*floare – floare***a** [floare] – [floarea] flower - the flower.

Romanian speakers have the intuition that the masculine definite article *-l*, as well as the final *-l* in *-ul*, are often not pronounced in the spoken language. However, studies on Romanian, generally focusing on the written version of the language do not provide any indication of how prevalent this phenomenon is in current spoken Romanian [8], [17].

The relatively frequent deletion of the definite article raised transcription questions during the data annotation process (production of the manual references). The presence of the masculine definite article seems to be strongly linked to type of speech: as can be expected, the more spontaneous the setting (talk shows, debates) the less carefully the definite article is pronounced.

From the 3.5 hour development corpus, 900 instances of words with word final *-ul* were manually selected and verified to avoid extraction errors. Then the data was segmented allowing pronunciations variants for *–ul* and *–u*. For this sample set, 20% of the variants preferred *-u* ending.

Inspecting the different instance, it was observed that the definite article is more likely to be **not pronounced** and **not transcribed**:

- before a following word that starts with a consonant as opposed to words starting with a vowel word (67% vs. 13%, respectively).

- in spontaneous and casual speech (debates) more than more formal, prepared continuous speech (broadcast news) (36% for a sub-corpus of debates vs. 20% for the entire dataset containing both data types.

In the 400+ hours of train data more then 100k occurrences of words with final *-ul* have been counted. The train and the dev data show similar trends, 12.6% of words ending in *-ul* preferred *-u* ending. This preliminary result confirms that automatic alignments can provide evidence of hypotheses observed in manual transcriptions [18].

5.2. Secondary palatalization

Secondary palatalization in Romanian is not an underlying characteristic as in Celtic and Slavic languages, but only results from operating in certain phonological and morphological contexts. Final palatalization is described as a secondary articulation and is more likely to be undetected for liquids, than for fricatives/affricates or stops [6]. According to the Romanian phonologist [19], the final secondary palatalization is a true characteristic of the Romanian language and

holds the (rather isolated position) that phonemic system possesses then a two series of consonants (similar to Russian): one palatalized and the other not.

The final palatalization concerns mainly plural nouns and adjectives and 2nd person singular indicative of verbs:

domn – domni [domn] – [domnj] mister – misters

sar – sari [sar] – [sarj] I jump – you jump

944 contextual occurrences with potential palatalization were automatically extracted from the 3.5 hour development data. The criteria was that the instance be one of: plural nouns, adjectives and 2nd person singular indicative of verbs. Based on the above observations, during segmentation the authorized pronunciation variants are: C (plain consonant alone), Ci (consonant + vowel i, i.e., true palatalization with Romanian model), Cj (palatalized consonant as in Russian, i.e., using the Russian palatalized consonants borrowed from Russian acoustic models).

Alignment of the corpus using the above variants gives the following preferences:

- C (undetected palatalization) 45.5%;

- Ci (detected palatalization with Romanian model) 32.3%;

- Russian Cj (true palatalization as in Russian) 20.2%.

These results highlight that word-final secondary palatalization in Romanian covers a range of contextual realizations.

6. CONCLUSIONS

This paper has reported on some of the challenges faced in developing an ASR system for the Romanian language, a low resourced language in the enlarged European Union, and on using the ASR system as a tool to study linguistic phenomena in modern spoken Romanian. Since only a small corpus of transcribed speech data were available for system development, the Romanian system was developed using text and audio data downloaded from the web. The acoustic models were training in a semi-supervised manner. Since many of the downloaded texts did not contain diacritic marks, these were automatically added to the text corpus. A typology of the main ASR transcription errors on the development corpus was made, which led to the study of pronunciation variants to capture two important linguistic phenomena: the deletion of the definite article and the presence vs. neutralization of the secondary palatalization of plural nouns and adjectives and of 2nd person singular indicative of verbs. Results on 3.5 hour corpus of manually transcribed speech show that the two phenomena are highly dependent of the surrounding context. However, the small amount of data does not allow us to assess if the variation is unpredictable or governed by specific rules (i.e., to sort out the "random variation" vs "language evolution" hypotheses). From the ASR standpoint, the study

and subsequent modeling of such pronunciation variants may help to reduce transcription errors. In future studies we plan to apply this type of methodology to a larger set of linguistic hypotheses, such as the general reduction of word final affixes in continuous speech. Since it is necessary to validate such hypotheses on large corpora, our aim is to validate results on both large corpora of automatically transcribed speech with observations on more limited manually transcribed data sets. Such studies can serve to contribute to linguistic knowledge and language evolution, as well as to improve the performance of automatic speech recognition systems.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] C. Burileanu H. Cucu, L. Besacier and A. Buzo, "Asr domain adaptation methods for low-resourced languages: Application to romanian language," in *European Signal Processing Conference*, 2012, pp. 1648–1652.

[2] C. Petrea D. Ghelmez-Hanes C. Burileanu, A. Buzo and H. Cucu, "Romanian spoken language resources and annotation for speaker independent spontaneous speech recognition," in *Conference on Digital Telecommunications*, 2010, pp. 7–10.

[3] L. Lamel and B. Vieru, "Development of a speech-to-text transcription system for finnish," in *SLTU*, Malaysia, 2010, pp. 62–67.

[4] M. Adda-Decker, "De la reconnaissance automatique de la parole a l'analyse linguistique des corpus oraux," in *Journaux d'Etude sur la Parole*, France, 2006.

[5] J.J. Ohala, "The connection between sound change and connected speech processes," pp. 201–206, 1996.

[6] I. Vogel L. Spinu and H.T. Bunnell, "Palatalization in romanian-accoustic properties and perception," *Journal of Phoenitics*, vol. (40)1, pp. 54–66, 2012.

[7] V. Barbu Mititelu D. Cristea D. Trandabat, E. Irimia and D. Tufis, "The romanian language in the digital age," in *META-NET White Paper Studies*, 2012.

[8] Th. Hristea, "Sinteze de limba română," *Editura Albatros*, 1984.

[9] G. Brâncuș, "Vocabularul autohton al limbii române (the autochtone vocabulary of romanian language)," *Editura*

Științifică și Tehnică (Scientific and Technical Publishing House), 1983.

[10] V. Niculae L.P. Dinu and M. Sulea, "Dealing with the grey sheep of the romanian gender system, the neuter," in *COLING*, Mumbay, 2012, pp. 119–124.

[11] L. Lamel J.L. Gauvain and G. Adda, "The LIMSI broadcast news transcription system," *Speech Communication*, vol. 37(1-2), pp. 89–108, 2002.

[12] J.L. Gauvain L. Lamel and G. Adda, "Lightly supervised and unsupervised acoustic model training," *Computer Speech and Language*, vol. 16, pp. 115–129, 2002.

[13] L. Lamel P. Fousek and J.L. Gauvain, "Transcribing broadcast data using mlp features," in *Interspeech*, Brisbane, 2008, pp. 1433–1436.

[14] J.L. Gauvain B. Vieru C. Woehrling V.B. Le J. Despres, L. Lamel and I. Oparin, "The vocapia research asr systems for evalita 2011," in *Evaluation of Natural Language and Speech Tools for Italian*, 2013, pp. 286–294.

[15] N. Snoeren M. Adda-Decker I. Vasilescu, D. Yahia and L. Lamel, "Crosslingual study of asr errors: on the role of the context in human perception of nearhomophones," in *Interspeech*, Firenze, 2011.

[16] N. Snoeren D. Yahia M. Adda-Decker, I. Vasilescu and L. Lamel, "Exploring linguistic variation in asr errors: Paradigm and tool for perceptual experiments," Philadelphia, 2011.

[17] I. Vintila-Rădulescu, "Dicționarul normativ al limbii române. ortografic, ortoepic, morfologic și practic," 2009.

[18] M. Adda-Decker M. Candea and L. Lamel, "Recent evolution of non standard consonantal variants in french broadcast news," in *Interspeech*, Lyon, 2013.

[19] E. Petrovici, "Sistemul fonematic al limbii române (the phonematic system of romanian)," 1956, vol. VII, f. 1-2, pp. 7–20.