THE PROSODY OF THE DOUBLE-SUBJECT SENTENCES IN ROMANIAN

HORIA-NICOLAI TEODORESCU, DIANA TRANDABĂȚ

Abstract. The double subject in Romanian sentences is a controversial linguistic phenomenon. While some researchers accept it as a language 'curiosity', others consider it apposition, in order to embody its behavior in the already existing theories. We present a first study in the literature on the phonetic analysis of double-subject sentences; the study is performed on spoken Romanian language.

1. INTRODUCTION

The purpose of this paper is to analyze the specificities of the prosody of the double-subject sentences. This topic has never been addressed, at our best knowledge, but the wider subject of linking prosody and other spoken language characteristics to the semantics of the oral message has been addressed by other authors too. For example, Daniel Hirst says "The way in which prosody contributes to meaning is still, today, a poorly understood process corresponding to a mapping between two levels of representation for neither of which there is any general consensus. It is argued that annotation of prosody generally consists in describing both prosodic function and prosodic form, but that it would be preferable to clearly distinguish the two levels." (Hirst, 2004). In fact, as Hirst emphasizes, "Everybody agrees that prosody contributes to the meaning of an utterance", yet there is little-known about how the prosody works at the phonetic level to enhance – or even change – the meaning expressed by the natural language words and phrases.

All European languages, as far as we know, use appositions to emphasize a specific meaning the speaker wishes to convey. Some languages, like the Japanese and the Korean languages, use for similar purposes specific constructions, named "double-subject constructions", but most modern languages, like English or French, do not use such constructions. In the Romanian linguistic community, there has been in recent years a debate on some types of sentences, which are considered by several researchers (Barbu, 2003, Cornilescu, 1997) and by us as being double-subject constructions.

After presenting the different approaches to double-subject sentences in Section 2, we explain the methodology behind the double-subject corpus creation and its analysis: annotation, acoustic parameters determination, etc. We present in

RRL, **LIII**, *4*, p. 517–528, Bucureşti, 2008

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

Section 4 the results of the prosodic analysis of double and simple subject sentences, before drawing some conclusions and indicating some further directions.

2. DOUBLE-SUBJECT SENTENCES IN ROMANIAN

The semantic arguments of a predicate (the subject, the direct object and the indirect object) can be doubled, in the Romanian language. While the objects are commonly doubled by clitic pronouns (the doubling is sometimes mandatory, like in ***L***-am văzut ***pe Ion***), the subjects receive, occasionally, and mainly colloquially, a doubling pronoun (not only in Romanian, as (Masahiro, 1996) shows[1]). The doubling of the subject for the Romanian language is a controversial phenomenon: after having long been considered an apposition, Alexandra Cornilescu (1997) has reopened the doubling problem, Verginica Barbu (2003) has modeled it using HPSG instruments, but until today, there is no unitary consensus. In this context, supplementary information should be gathered on the specificities of the double-subject constructions contrasted both to the single subject sentences and to sentences that include appositions. Specific phonetic constructions for the three cases would be a significant argument for three independent linguistic constructions. What supplementary information the pronouncing brings, from a descriptive perspective, in double-subject phrases, remains an open question. The present paper partially answers this question.

Some examples of sentences with double subject are:

(a) Vine ea mama!

(b) "A trecut el aşa un răstimp" (Sadoveanu M.)

A principle we wish to introduce and use here is that consistent distinctions at the phonetic level between two specific constructions reflect and represent an argument to distinguish at the syntactical level between the two constructions.

The first author proposes that the double-subject sentences convey different meanings, depending on the prosody, for example:

− a neutral pronunciation indicates a non-determination of the time interval.

− a pronunciation accentuating the pronoun "el" indicates that the speaker has an idea about the time interval duration, and that the focus is on the passing of that time, and not on the duration.

− if the sentence is further developed, it can bring a further specification of the interval. For example, in the development „A trecut el aşa un răstimp de lung, încât...", the duration of the interval is specified in a certain way.

(c) O şti el careva cum să rezolve asta.

[1] There is no definite explanation why not all languages accept the double-subject structure. For these languages, in most of the cases, the doubling of the subject is realized as an apposition. Romanian language uses both double subject and apposition structures.

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

Different pronunciations may mark the fact that the speaker does not know who is the person mentioned ("el"), or that he knows, but has no intention on telling to the audience (when the accent is on „careva"), or clearly specifies, by an apposition, who is envisaged – if the sentence is developed (as in „O şti el careva, Ionică, cum să rezolve asta".) Notice that such a sentence, including both apposition and double subject, is a strong argument in favor of the existence of the double subject constructions.

For the examples b) and c), the interpretation is that the information must be partially known by the auditorium (knowledge at the generic level, but not at the level of instantiation with a concrete individuality).

(d) Mama vine şi ea mai târziu.

(e) Mama ştie ea ce face.

Some linguists (Barbu, 2003), considered cases like in examples d) and e) as constructions with doubled subject, while other authors (Cornilescu, 1997) consider them particular structures of the Romanian language. We intend to compare them to see if there are differences in their prosodic realizations.

In this context, we recorded a set of sentences bearing doubled subject for a comparative analysis of the prosody in sentences with doubled- and simple-subject. We performed the recordings with the aim to determine the modifications – if any – involved in the prosody by the doubling of the subject. The main objectives of our study are, specifically:

− To compare the prosody for simple subject and double-subject sentences;

− To clarify the prosodic aspects and differences, if any, between the standard double-subject constructions (examples (a)-(c)) and the non-standard structures (examples (d) and (e));

− To study the modifications induced by the doubling of the subject in the sentence prosody;

− To correlate the semantic charge with the pronunciation (different accentuation of the sentences with doubled subject);

− To determine if the spoken language brings distinctions that may change the sentence behavior closer to a simple subject construction or a double subject one.

While the hypotheses stated above are not yet statistically validated, in this paper we bring clarifications on the change of prosody in double-subject sentences in comparison with simple sentences.

3. METHODOLOGY

In order to realize a correlation between the semantic charge carried by a sentence and the representation of its subject, we have recorded, for several speakers, the five sentences presented in Section 2. The speakers are aged between 26 and 31 years, born and educated in the Middle Moldavian region (counties Iaşi, Vaslui, Bacău); all have university education and they have no manifested pathologies.

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

The recordings were included in the Romanian Sounds Archive (Romanian Sounds Archive), where they are freely accessible. The archive contains over 800 distinct recordings, some of them (soon all of them) available in various accuracy and encoding formats. Apart the archive itself, the site hosts also documentations regarding the description of the technical modalities and conditions (protocols) involved by the realization of the archive. Namely, the database contains two types of protocols:

− The documentation protocol, which contains the speaker profile (linguistic, ethnic, medical, educational, professional information about the speaker), and a questionnaire regarding the speaker's health, especially concerning the pathologies of the phonating tract.

− The recording protocol, containing information about the noise acceptable values, the microphone, the soundboard, and the corresponded drivers.

3.1. A double-sentence spoken database

The subjects have been informed about the objectives of the project; they signed an informed consent according to the Protection of Human Subjects Protocol of the U.S. Food and Drug Administration and to the Ethical Principles of the Acoustical Society of America for Research Involving Human Subjects. The speakers' selection was tributary to the Archive's constraints (the documentation protocol).

The recordings (sound files) corresponding to the simple subject and double-subject sentences have been recorded according to the methodology explained in the recording protocol of the Romanian Sound Archive (Romanian Sounds Archive). The recordings have been performed using the GoldWave™ application, with a sampling frequency of 22050 Hz. The accuracy is given by the number of bits per sample, and in this regard we offer two versions: 16 bits and 24 bits per sample. The available encodings are *.wav*, *.ogg* (free codec, used by the international community), and *.txt* (ASCII encoding, universal) file types.

The speakers[2] have recorded several variants of the five sentences mentioned in Section 2; the sentences have been uttered with neutral tone, accentuation of the doubling pronouns, focuses on the words next the pronouns, or the extension of the sentences.

(a) Vine ea mama!
(b) "A trecut el aşa un răstimp" (Sadoveanu)
(c) O şti el careva cum să rezolve asta.
(d) Mama vine şi ea mai târziu.
(e) Mama ştie ea ce face.

[2] Several speakers have been recorded for the double/simple subject analysis. The results discussed in Section 4 consider five subjects: subject #1 (female), subject #2 (female), subject #5 (male), subject #7 (male) and subject #12 (female).

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

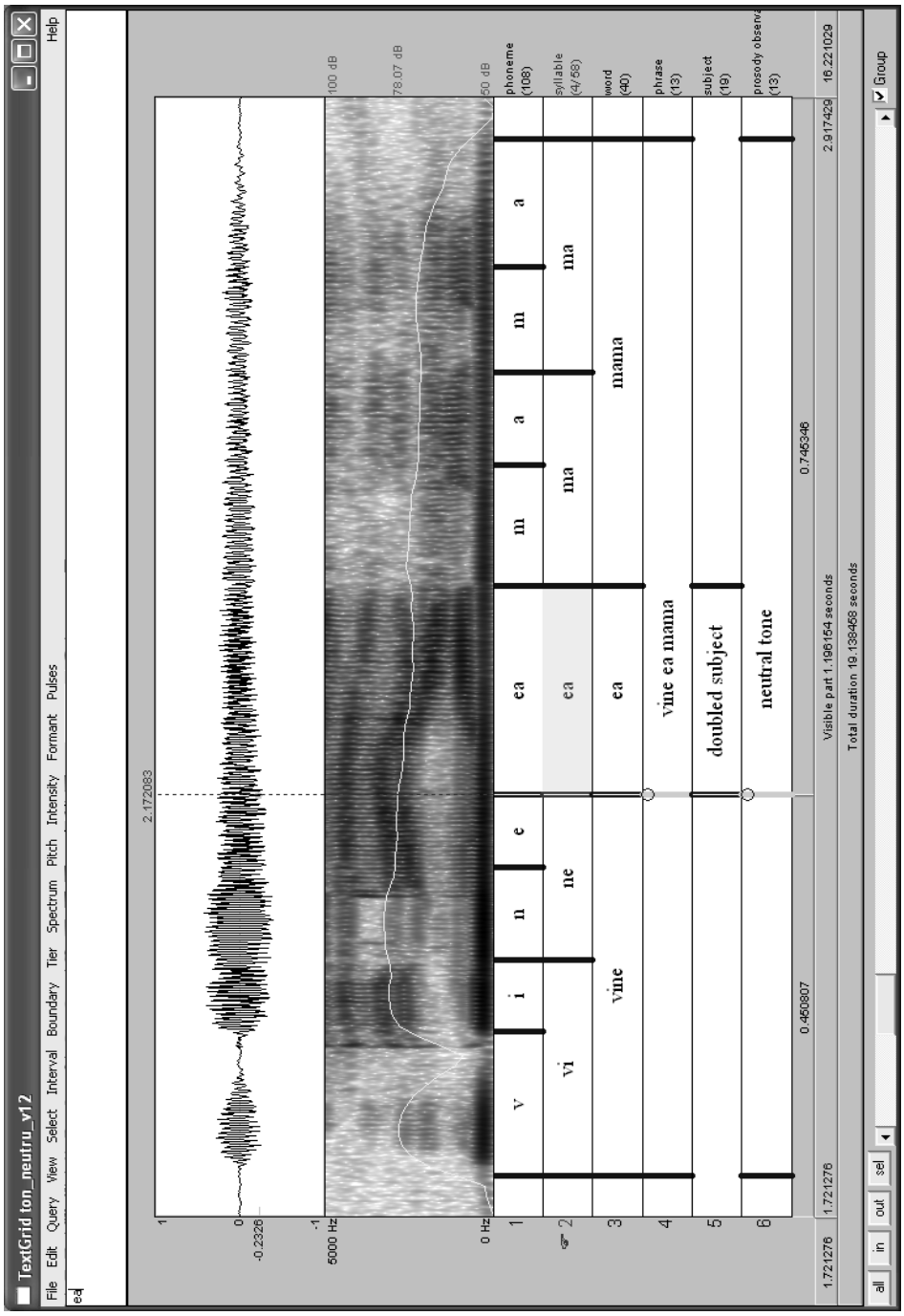


Fig. 1 – Example of annotation using Praat™ of the sentence “Vine ea mama”.

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

Also, the corresponding sentences without the double subject have been recorded. Every speaker pronounced each sentence three times, following the archive recording protocol (see, for details, (Romanian Sounds Archive)).

3.2. Analysis methodology

The analysis of the double subject requires finding and correlating the double sentences parameters with the corresponding simple sentences parameters. The sentences have been annotated using the Praat™ software (Praat) at several levels: phoneme, syllable, word, sentence, subject position, and articulation type. After the annotation, the formants are determined for the sentence vowels and semi-vowels. For an as precise as possible determination, we have selected segments of the vowels fulfilling the following conditions:

− The selected segment should be a central area, where there are no transitions of the formants to those of the joined phonemes;

− The formant's frequency should not present large fluctuations. The fluctuations of the formants and their correlation to the double subject will be analyzed as a subsequent step;

− The formant's contour should not contain interruptions.

An example of an annotated sentence is presented in Fig. 1.

Unfortunately, different analysis tools provide different results. This is partly because there is no single definition for the parameters of the non-stationary signals (as the speech signal is), various tools using different *ad hoc* definitions. Therefore, we have applied several programs, namely Praat™ (Praat), Klatt analyzer™ (Klatt), GoldWave™ (GoldWave) and WASP™ (WASP) to determine the acoustic parameters. We discuss the obtained results in the next Section.

4. DOUBLE-SUBJECT SENTENCES ANALYSIS

The hypothesis that motivated this analysis is that the double-subject constructions relate in a specific way to the emotional and to the inter-relationship representation. We contrasted therefore the values of the formants and duration of the vowels for five subjects (three female and two male) from our database for the sentence "`Vine mama`" (simple subject) vs. "`Vine ea mama`" (doubled subject). We realize that an analysis over five subjects can have no claims on generality, but it represents a good start for the phonetic analysis of the Romanian double subject constructions. The recorded subjects belong to the same age bin (25-30 years), have higher education and came from the same geographic area.

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

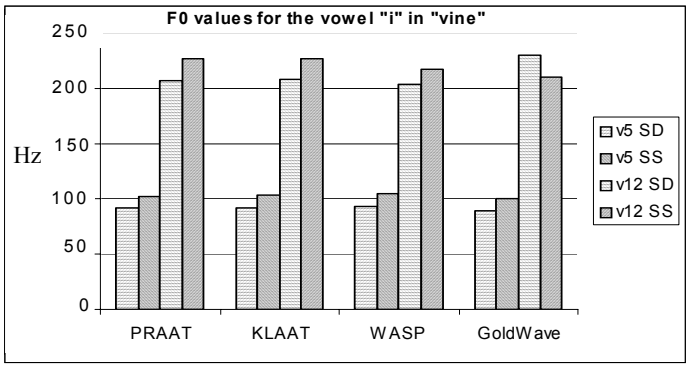


Fig. 2a – F0 evolution for “i” in “Vine mama”.

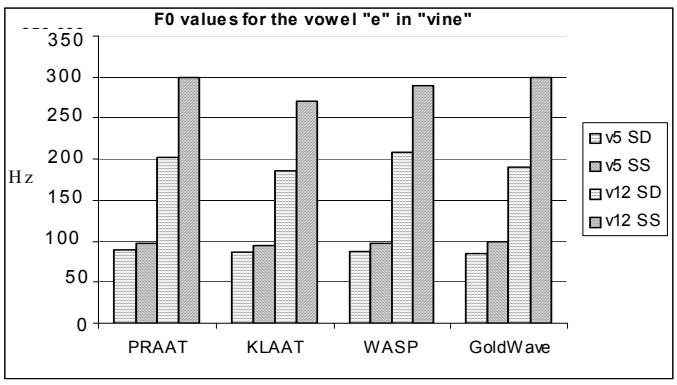


Fig. 2b – F0 evolution for “e” in “Vine mama”.

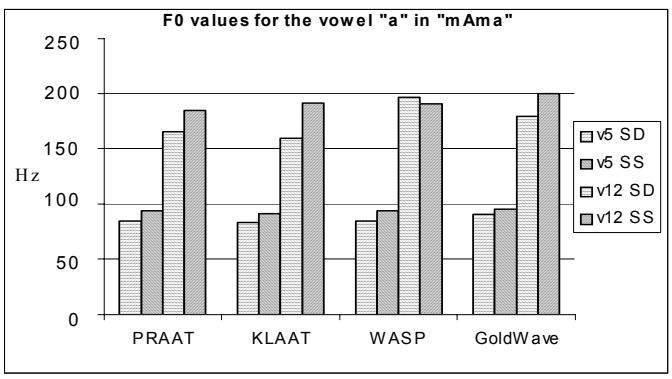


Fig. 2c – F0 evolution for “a1” in “Vine mama”.

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

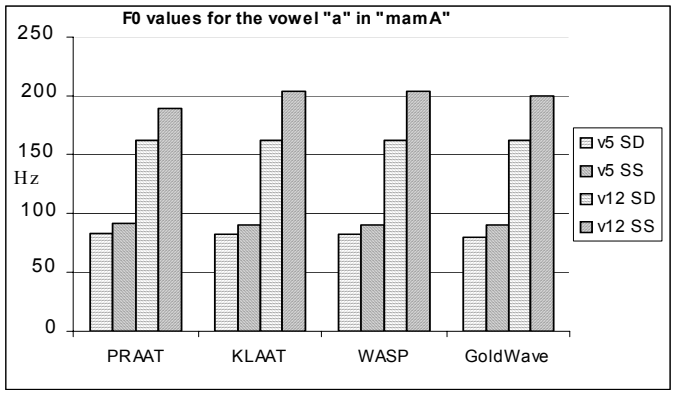


Fig. 2d – F0 evolution for "a2" in "Vine mama".

The results of the analysis are graphically summarized in the accompanying figures. *Figure 2* presents the F0 values obtained with the four considered analyzers for two subjects: subject #5 – male and subject #12 – female. In the legend, SD stands for "double subject" and SS for "simple subject".

In the graphs, the first two bars – for each analysis program – represent the values for the male subject (double-subject sentence vs. simple subject sentence), while the last two are the F0 values for the female subject. When looking at the F0 values for the vowels of the analyzed sentence (namely the vowel ***i*** in Fig. 2a, ***e*** in Fig. 2b, the first ***a*** in *mama* in Fig. 2c and the second ***a*** in Fig. 2d), we noticed that all the four programs show an increasing tendency of the F0 values for all the vowels in the simple subject sentences vs. double-subject sentences.

Using the values presented in Fig. 2, we have computed a mean value for the values of the first four formants obtained with different analysis programs. Then, in order to see how significant the increasing of the pitch is, we computed also the standard deviation of the four values relative to the average. The obtained values are presented in Table 1.

Table 1

The mean for the F0 and the standard deviation, in Hz

	Vine ea mama					Vine mama			
	i	e	mAma	Mama		i	e	mAma	mama
	F0	F0	F0	F0		F0	F0	F0	F0
subject #12 ST DEV	12	10	17	0		8	14	6	7
subject #12 MEAN	212	196	175	161		220	290	192	199
subject #5 ST DEV	1	2	3	1		2	2	1	0
subject #5 MEAN	92	87	85	82		102	98	94	91

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

We observed that the major differences in the mean pitch values are visible for the vowels in unaccentuated syllables. Thus, the pitch of the vowel ***e*** in *vine* decreases in the simple subject structures by an average of 93.53 Hz for subject #12, and by an average of 10.27 Hz for subject #5, while the last ***a*** in *mama* decreases by 36.84 Hz for the female subject and by 8.81 Hz for the male speaker. In the accentuated vowel case, the decreasing is lower (on average, by 8.28 Hz for the ***i*** in *vine* for subject #12, and 10.13 Hz for subject #5, and respectively 16.65 Hz for the first ***a*** in mama for subject #12, and 8.35 Hz for subject #5). A possible cause for this changes, that deserves a more detailed analysis may be the location of the unaccentuated vowels at the end of the words.

The growing tendency of the F0 values is obvious also for the other subjects. For the same sentences, the mean values obtained for the pitch, for the vowel ***a,*** are shown in Table 2.

Table 2

Values for the vowels in the subject of "`Vine mama`" vs. `Vine ea mama`"; frequency in Hz and time in s

	Vine ea mama				**Vine mama**			
	a1 in mAma		**a2 in mama**		**a1 in mAma**		**A2 in mamA**	
	F0	duration	F0	Duration	F0	duration	F0	duration
subject #1	200	0.086	215	0.082	211	0.103	223	0.098
subject #2	189	0.101	179	0.137	215	0.067	206	0.098
subject #12	162	0.099	162	0.135	188	0.127	196	0.136
subject #5	84	0.094	83	0.084	93	0.122	91	0.138
subject #7	76	0.080	71	0.079	77	0.089	82	0.070

The data recordings we have annotated are not sufficient to draw statistically pertinent conclusions for the vowels duration changes. For now, we can only say that the tendency to increase or to decrease the duration of the vowels seems similar in both construction types. Thus, while, for example, in the double-subject construction, subject #12 has increased the duration of the last vowel *a*, this increasing tendency is also found in the simple subject structure. The effect is similar for subject #1, but with decreasing tendency. However, subject #5 disobeys this rule, while the values obtained for subject #7 are too close to be considered relevant.

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

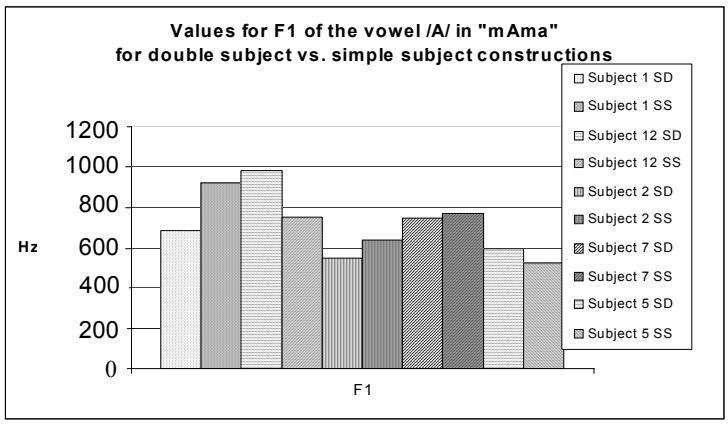


Fig. 3a – F1 evolution for "a1" in "Vine mama".

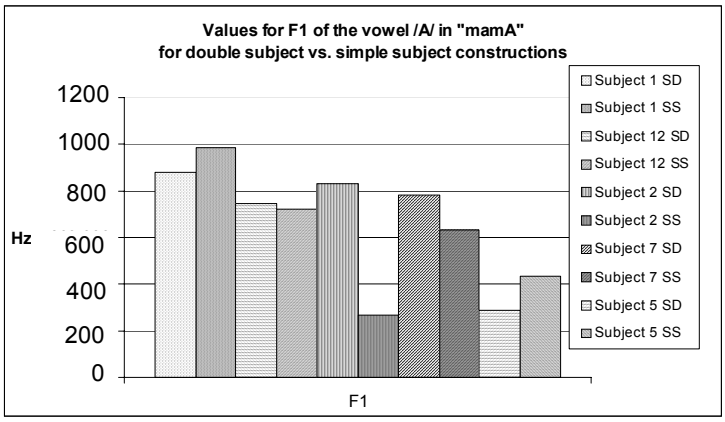


Fig. 3b – F1 evolution for "a2" in "Vine mama".

The values for the first formant are presented in Fig. 3. Notice that, for the first *a* vowels in the sentence (Fig. 3a), three subjects have increased their F1 values, while two have decreased them. For the second *a* (Fig. 3b), there is an inverse tendency: three values decrease, while two increase. It looks that the first formant is fluctuating and carries no double subject information. However, it carries information about the speaker. We can therefore make no generalizations, until more subjects are considered. However, we may notice that the increasing / decreasing tendency is kept by the speaker for the two vowels, with the exception of the last two speakers (male). We have to validate this exception through further analysis.

For the rest of the formants, the values show no regularities. We envisage therefore two directions for further analysis of the superior formants: one involves collecting more data and continuing the presented approach, the other intends to

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

use a decision instrument to observe if there are regularities that have not been noticed by human annotators.

After analyzing several double / simple subject constructions, we believe that the hypothesis we have started with is at least partly proved. The inter-relationship between the speaker and its audience becomes visible by the observation that the speaker has already a pattern (referring to the pitch contour) when beginning to pronounce a structure (higher pitch for simple subject structures, lower values for double subject).

5. CONCLUSIONS AND FURTHER WORK

We have analyzed the influence of the double-subject construction on the prosody in the Romanian language. The analysis involved short sentences that are parallel in the sense that they are identical up to the use of single or double-subject constructions.

The main conclusion derived from this preliminary research is that the pitch differs in a consistent way between the two types of sentences. Namely, the pitch is lower in double-subject constructions than in single subject sentences. A second conclusion is that the frequency of the first formant changes between the two constructions, but the way of changing and the change amplitude depend significantly on the speaker. The vowel duration also may change, but there is no a single type of change; however, while the inter-subject changes are inconsistent, the intra-subject change tend to be consistent. These findings tend to support the idea that the constructions some researchers argue to be double-subject sentences are different, prosodically and possibly semantically, compared to single subject constructions.

Future analysis will be devoted to contrast the prosody of parallel sentences with double-subject constructions, appositions, and simple (i.e., no apposition, no double subject) constructions in the Romanian language.

Acknowledgments. We acknowledge the partial support of the Romanian Academy "Priority Research" program "Sisteme cognitive". We thank Prof. Corneliu Burileanu and his colleagues for valuables comments.

REFERENCES

Barbu, V., 2003, "Construcţii cu subiect dublu în limba română actuală. O perspectivă HPSG", in: G. Pană Dindelegan (coord.), *Aspecte ale dinamicii limbii române actuale,* vol. II, Bucureşti, Editura Universităţii din Bucuresti, 73−79.

Cornilescu, A., 1997, "The Double Subject Construction in Romanian. Notes on the Syntax of the Subject", *Revue roumaine de linguistique*, 3−4, 1−45.

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)

Hirst, D., 2004, "The Phonology and Phonetics of Speech Prosody: Between Acoustics and Interpretation", *Speech Prosody,* Nara, Japan, March 23-26, 2004. ISCA Archive, http://www.isca-speech.org/archive (accessed 21 Jan. 2007)

Praat: doing phonetics by computer, http://www.praat.org (accessed 2 March 2007)

(Klatt) *KPE80 − A Klatt Synthesiser and Parameter Editor*, http://www.speech.cs.cmu.edu/comp.speech/Section5/Synth/klatt.kpe80.html (accessed 2 March 2007)

GoldWave – Audio software, http://www.goldwave.com (accessed 12 Nov. 2006)

(WASP) *UCL Phonetics & Linguistics, WASP – Waveform Annotations Spectrograms and Pitch*, http://www.phon.ucl.ac.uk/resource/sfs/wasp.htm (accessed 2 March 2007)

(Romanian Sounds Archive) http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.htm (accessed 2 March 2007)

Masahiro, O., 1996, "Analyzing Japanese double-subject construction having an adjective predicate", *Proceedings of the 16th conference on Computational linguistics*, Volume 2, Copenhagen, Denmark, 865–870.

BDD-A299 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 19:59:30 UTC)