INFORMATICA, 2001, Vol. 12, No. 2, 315–336

Automatic Stressing of the Lithuanian Nouns and Adjectives on the Basis of Rules

Pijus KASPARAITIS *
Department of Informatics, Faculty of Mathematics and Informatics, Vilnius University
Naugarduko 24, 2006 Vilnius, Lithuania
e-mail: pkasparaitis@yahoo.com

Received: August 2000

Abstract. The paper deals with automatic stressing of the Lithuanian text. In the previous work the author presented an algorithm for automatic stressing of the Lithuanian text on the basis of a dictionary. The aim of the present work is to improve the above mentioned algorithm by including formal stressing rules for nouns and adjectives. By means of these rules such words as diminutives, names and degrees of adjectives that are not present in the dictionary may be stressed. The work analyses when it is more convenient to formulate rules manually and when to generate them automatically. A method for formulating rules manually has been described and a set of such rules has been presented. Besides the algorithm for generating stressing rules with the help of a dictionary of stems of nouns and adjectives has been given.

Key words: automatic text stressing, formal rules.

1. Introduction

The algorithm enabling to stress a Lithuanian text using dictionaries of stems was presented in (Kasparaitis, 2000). The largest dictionary was that of nouns and adjectives (containing over 53000 stems), therefore it will form the subject of the analysis.

It should be noted that some words having identical endings of the stem (infixes) are stressed the same way. E.g., the words "filolo`gas", "geolo`gas", "psicholo`gas" etc. are stressed by means of the first stressing paradigm, the words "brole∼lis", "name∼lis", "vaike∼lis" etc. – by means of the second one. These examples enable us to formulate the following rules: the words that match the template "*logas" are to be stressed by means of the first stressing paradigm, whereas the words that match the template "*elis" – by means of the second one, where the character '*' denotes any string of characters.

It seems possible to formulate rules generalising stressing of the whole group of words and thus enabling the following aspects of the algorithm to be improved:

1) to reduce the volume of the dictionary;

2) to stress diminutives and words with infixes. A current dictionary contains only those words the meaning of which is changed by adding the infix. To stress all words

*This research was partially supported by Lithuanian State Science and Studies Foundation.

with infixes, they have to be stored in a dictionary too, which would lead to an additional increase in the volume of a dictionary. Since words in the Lithuanian language can have an unlimited number of infixes (e.g., “sūn-ait-ėl-išk-as”), it is simply impossible to have all possible variants of a word with infixes in a dictionary;

3) to stress unknown words by means of word building rules, e.g., the surname “Kaspara′itis” is stressed like the word ”našla′itis” and ”Dudė′nas” – like ”meškė′nas” (no surnames and other proper nouns were included in the dictionary at all in the above mentioned paper);

4) to simplify stressing of the degrees of adjectives. In the previous algorithm the infixes of the comparative and superlative degrees of adjectives were treated as parts of endings, which made the algorithm bulky and difficult to understand;

5) to stress the reflexive nouns – abstracts of the action, e.g., “neši`masis”.

This paper discusses what requirements should satisfy such rules, what form the rules should have, the way of formulating, storing and using the said rules in the search.

2. Requirements for Rules

The rules should satisfy the following requirements:

1) the rules and the stems should be stored in the same dictionary;

2) having replaced the groups of identically stressed stems by the rules, the number of records in the dictionary (stems and rules) should be minimal;

3) the use of stressing rules should not lead to any additional errors or bring about polysemy in the stressing results as compared with the use of the whole stems. E.g., if the rule “*ė′ja” were formulated to stress the nouns “siuvė′ja”, “mezgė′ja”, “šlavė′ja” etc., it would also be used (erroneously) to stress the verb “eižė′ja” (correct stressing is “e′ižėja”);

4) the rules should enable us to stress grammatical forms with infixes that are not usually stored in the dictionary: diminutives (e.g., “name∼lis”), the degrees of adjectives (e.g., ”gere`snis”, ”gerėle`snis”, ”geria′usias”);

5) the rules should enable us to stress words that have a somewhat more complicated structure, e.g., words that are build using several infixes (“nameliu`kas”, “namukė∼lis”).

3. Form of Rules and Restrictions

The rules that are similar in their form to the records stored in the dictionary (Kasparaitis, 2000) for each stem will be used. The following record is stored in the dictionary for each stem: the stem, the type of stem, the stressing paradigm, the stress position and the accent. An element consisting of a string of letters and the ending of the stem will be used in the rules instead of the stem. The record containing the whole stem can be treated as a record of the rule where the stem corresponds to the ending of the stem and the string of letters is empty. Since the records of rules and the records of stems have the same form, so they can be stored in the same dictionary.

A note. To make the rules more understandable the following simplified marking of records of rules and stems will be further used: the whole rule will be given between the quotation marks, an asterisk '*' will denote a string of letters which will be proceeded by the restrictions applied to it. Further, the ending of the stem is written in the lowercase letters. A dash and the ending of the singular nominative case in uppercase letters are written instead of the number of type of the stem (if this does not lead to polysemy). The type of the stem defines a set of endings that can be added to a certain stem, e.g., the word "namas" has the type of the stem "-AS"={"-as", "-o", "-ui", "-ą", "-u", "-e", "-e", "-ai", "-ų", "-ams", "-us", "-ais", "-uose"}. Further, the number of the stressing paradigm separated by the underscore is written. Stressing paradigms are defined in (Ambrazas, 1996). The place of the stress and accent are defined by putting the sign of a stress on the corresponding letter of the stem (or the ending of a stem). In some examples the sign of a stress is placed on a string of letters. In these cases every omitted letter is denoted by the question mark. Thus, the above mentioned rules can be written as follows:

"*lo`g-AS_1";

"*e∼l-IS_2".

Different restrictions can be applied to a string of letters. E.g., a string of letters must have at least two syllables in the first rule presented above, whereas in the second one – it must consist of the stem of a noun of the masculine gender, i.e., a stem of a word with a corresponding type of the stem must exist in the dictionary of nouns and adjectives (or a matching rule must exist in the set of rules, if the dictionary of nouns and adjectives was replaced with the set of rules). Such restrictions ensure that the rules should not be applied to the words to which they must not be applied. E.g., the first rule must not be applied to the adjective "blo∼gas", and the second one – to the noun "vie∼škelis". Consequently, some kinds of strings of letters will be used depending on the restrictions.

Words that are generalised by one rule must have the same type of a stem, the stressing paradigm, the stress position and the accent. Besides, a certain number of letters in the ending of a stem must coincide. Only the rest of the stem (a string of letters) can be different.

In the rules being formulated the stress can be put:

1) on the ending of a stem, e.g., "*e∼l-IS_2";

2) on a string of letters, e.g., the rule "*išk-AS_1" for the word "že∼miškas".

In the second case the stress position is not defined in the rule, however, a string of letters must satisfy the additional restriction, i.e., it must define the stress position and the accent. The rules that define the stress position outside the ending of the stem can not be used. E.g., if the rule "*∼?išk-AS_1" were used for the word "že∼miškas", this rule would be erroneously applied to the word "pu`sininkiškas".

The ending of the stem in one rule can be a subset of the ending of a stem in another rule and these rules can define a different way of stressing. E.g., the following rules are possible: "*a∼lin-IS_1" ("genera∼linis") and "*i`n-IS_2" ("nami`nis"). It is convenient to assign certain priorities to the rules and to apply merely the rule with a higher priority. A longer rule usually has a higher priority. In case two rules are of the same length, both are applied, e.g., "py′lim-AS_1" and "pyli`m-AS_2".

A string of letters of one rule together with the ending of a stem can be used as a string of letters of another rule. Thus, a recurrent sequence of rules enabling us to model words with many infixes is obtained. E.g., with the rules "D*e∼l-IS_2" and "D*iu`k-AS_2" available, where 'D*' denotes the stem of a noun, the word "nam-el-iu`k-as" can be stressed. Not all the rules can build recurrent sequences. E.g., the rules "B*e`sn-IS_4" and "B*ia′us-IAS_1" (where 'B*' is a stem of an adjective) that model the comparative and superlative degrees of adjectives are not used for this purpose because no such words as "ger-esn-iaus-ias" or "ger-iaus-esn-is" exist in the Lithuanian language. Such rules can only be used at the right end of the recurrent sequence of rules. If we have the rule "*im-AS_3", we can stress such words as "art-im-e`sn-is" or "art-im-ia′us-ias".

Endings of stems of some rules can begin with the letter 'i' that performs a role of the sign of softening. In this case assimilation can take place before the ending of the stem, i.e., the letter 't' changes to 'č' and 'd' to 'dž' (analogously as in the case of endings, which begin with the letter 'i' that performs a role of the sign of softening). If a string of letters must be a stem of a word or something like that, assimilation should be taken into account before making the search for a stem.

4. Definition of Compatibility of Rules

If in formulating rules words were simply joined in groups without taking into consideration the rules that have already been formulated, it might so happen that several rules defining a different stressing could be applied to the word which is stressed in one way only. To avoid this happening one must formulate only compatible rules.

DEFINITION. Non-compatible rules shall be rules that have identical endings of the stem, restrictions of strings of letters allow us to have the same strings of letters, sets of endings defined by the type of the stem have at least one identical ending and the stress position or the accent differ in the cases defined by these identical endings.

Examples of non-compatible rules:

1) rules that have the same type of the stem, the stress position and the accent but different stressing paradigms: "*an-A_1" ("lie∼kana") and "*an-A_3" ("dovana`");

2) rules that have the same type of the stem and the stressing paradigm, however, a different stress position or the accent: "*o∼n-AS_2" ("baro∼nas") and "*o`n-AS_2" ("vago`nas");

3) rules that have different types of the stem, however, the endings in the genitive case plural are the same and they are stressed differently: "*at-IS_3" ("gaišati`s – gaišačių∼") and "*a∼t-Ė_2" ("kandida∼tė – kandida∼čių");

4) rules that have different types of the stem, but the ending of the nominative case singular in one rule is the same as the ending of the accusative case plural in the second rule and these cases are stressed differently: "*e`n-AS_2" ("šate`nas") and "*e`n-A_2" ("sirena`– sirena`s").

The last example provides the best illustration of how difficult it is sometimes to recognise non-compatible rules.

DEFINITION. Compatible rules shall be rules that either have different endings of the stem or restrictions of a string of letters permit having only different strings of letters, or sets of endings defined by the type of the stem have no identical endings, and in case they have, the stress position and the accent are identical in the cases corresponding to those endings.

Examples of compatible rules:

1) rules that have different endings of the stem: "*o′n-AS_1" ("lavo′nas") and "*ū∼n-AS_2" ("malū∼nas");

2) rules where restrictions of a string of letters permit to separate sets of strings of letters: "D1*inink-AS_1" ("mo′kslininkas") and "D3*inin∼k-AS_2" ("darbinin∼kas") where 'D1*' denotes the stem of the noun which is stressed using stressing paradigm 1 or 2 and 'D3*' – using stressing paradigm 3 or 4;

3) rules where types of stems define sets of endings that have no identical endings: "*i′en-A_1" ("avi′ena") and "*ie∼n-Ė_2" ("uogie∼nė");

4) rules when cases with identical endings are stressed in the same way: "*ai∼n-IS_2" ("riestai∼nis – riestai∼nių") and "*ai∼n-Ė_2" ("mišrai∼nė – mišrai∼nių");

Let us assume that a certain set of rules has been created (e.g., we have generated them automatically) containing some non-compatible rules. Non-compatible rules can be changed to compatible ones in the following way:

1) by expanding the ending of the stem of one rule (see the examples with the words "nam-i`n-is" and "gener-a∼lin-is" above);

2) by choosing suitable restrictions for strings of letters (see the examples with the words "mo′ksl-inink-as" and "darb-inin∼k-as" above).

What should be done providing non-compatible rules are impossible to be made compatible? To leave one rule unchanged and to enter all the stems that have been generalised by the second rule in the dictionary. Usually the rule that generalises more stems is left unchanged.

5. Investigation into Compatibility of Rules

It is often difficult to formulate strict restrictions that allow us to have only different strings of letters and in this way to make the rules compatible. Therefore it worth analysing the compatibility of rules separately without taking into account the restrictions.

Let us assume that pairs of rules that have the same endings of the stem and the same restrictions of a string of letters are being analysed. Let us make clear what combinations of the type of a stem and the stressing paradigm produce compatible rules by pairs, and whether it is necessary for the stress position and the accent to be identical to make the rules compatible.

The following algorithm was used to check the rules:

1) The first ending in the set of endings defined by the type of a stem is taken and it is defined whether all cases having that ending are stressed the same way. E.g. the word "namas" has the ending "-e" in two differently stressed cases: the locative case singular "name`" and the vocative case singular "na∼me".

2) All the cases with the above mentioned ending are found in the second rule. The rules are non-compatible if both rules have the stress in the stem, however, the stress position or the accent differs. If all the cases of the first rule that have that ending are stressed in the same way, the rules are non-compatible in one more case, i.e., when one rule has the stress in the ending and another rules – in the stem.

3) Step 1 and 2 are repeated for all the endings of the first rule. Rules are non-compatible if they are non-compatible for at least one ending.

4) The said procedure is applied once again after the rules have been exchanged because one rule may not bring any polysemy in another rule, whereas another rule may.

This algorithm was realised in the form of a computer program. As only stressing paradigms 1 and 3 and stressing paradigms 2 and 4 can have the same stress position and

Table 1

Compatibility of rules that have stressing paradigms 1 and 3

namas	brolis	arklys	ranka	vyšnia	bitė	krosnis	žvėris	sodžius	geras	žalias	didelis	medinis	gražus	Example defining the type of stem	
1 3	**1**	**3**	**1 3**	**1 3**	**1 3**	**1 3**	**3**	**1**	**1 3**	**1 3**	**3**	**1**	**3**	str. paradigm	
1 2 2 1	0 0	0 0	1 2 2 2	0 0 0 0	1 1 2 2	1 2 2 0	2 0	0 0	1 2 2 2	0 0 0 0	1 2	1 2	2 2	**1** **3**	namas
	1	2	0 0	1 2	1 2	1 2	2	1	1 2	1 2	2	1	2	**1**	brolis
		1	0 0	2 2	2 0	2 2	2	2	0 0	2 2	2	2	2	**3**	arklys
			1 2 2 1	0 0 0 0	0 0 0 0	1 2 2 0	2 0	0 0	1 2 2 2	1 1 2 2	0 0	0 0	1 2	**1** **3**	ranka
				1 2 2 1	1 2 2 0	1 2 2 1	2 0	1 2	0 0 0 0	1 2 2 2	2 2	1 2	2 2	**1** **3**	vyšnia
					1 2 2 1	1 2 2 0	2 0	1 2	0 0 0 0	1 2 2 0	2 2	1 2	2 0	**1** **3**	bitė
						1 2 2 1	2 1	1 2	1 2 2 0	1 2 2 2	2 2	1 2	2 2	**1** **3**	krosnis
							1	2	2 0	2 0	2	2	2	**3**	žvėris
								1	0 0	1 2	2	1	2	**1**	sodžius
									1 2 2 1	1 2 2 2	2 0	1 2	2 2	**1** **3**	geras
										1 2 2 1	2 2	1 2	2 2	**1** **3**	žalias
											1	2	2	**3**	didelis
												1	2	**1**	medinis
													1	**3**	gražus

Table 2

Compatibility of rules that have stressing paradigms 2 and 4

namas	brolis	arklys	ranka	vyšnia	bitė	krosnis	žvėris	sūnus	sodžius	geras	žalias	didelis	medinis	gražus	Example defining the type of stem	
2 4	**2**	**4**	**2 4**	**2 4**	**2 4**	**4**	**4**	**2 4**	**2**	**4**	**4**	**4**	**2**	**4**	**str. paradigm**	
1 2 2 1	0 0	0 0	2 2 2 2	0 0 0 0	1 1 1 1	2 0	2 0	2 2 2 2	0 0	2 2	0 0	2 2	1 1	1 2	**2** **4**	namas
	1	2	0 0	1 2	1 2	2	2	0 0	2	2	2	2	1	2	**2**	brolis
		1	0 0	2 2	2 0	2	2	0 0	2	0	2	2	2	2	**4**	arklys
			1 2 2 1	0 0 0 0	0 0 0 0	2 0	2 0	1 2 2 2	0 0	2 2	2 2	0 0	0 0	1 2	**2** **4**	ranka
				1 2 2 1	1 2 2 0	2 1	2 0	0 0 0 0	1 2	0 0	2 2	2 2	1 2	2 2	**2** **4**	vyšnia
					1 2 2 1	2 0	2 0	0 0 0 0	1 2	0 0	2 0	2 2	1 2	2 0	**2** **4**	bitė
						1	1	2 2	2	0	2	2	2	2	**4**	krosnis
							1	2 2	2	0	0	2	2	2	**4**	žvėris
								1 2 2 1	0 0	2 2	0 0	0 0	0 0	2 2	**2** **4**	sūnus
									1	0	2	2	2	2	**2**	sodžius
										1	2	0	2	2	**4**	geras
											1	2	2	2	**4**	žalias
												1	2	2	**4**	didelis
													1	2	**2**	medinis
														1	**4**	gražus

accent, three cases were investigated separately:

1) compatibility of the rules that realise stressing paradigms 1 and 3 by separating the cases when the rules have the same stress position and accent and when they have different ones;

2) compatibility of the rules that realise stressing paradigms 2 and 4 by separating the cases when the rules have the same stress position and accent and when they have different ones;

3) compatibility of the rules that realise stressing paradigms 1, 3 and those that realise stressing paradigms 2, 4. The rules have different stress positions or accents.

Results of experiments are presented in Tables 1–3.

The type of a stem is defined by a typical example of the word that has this type of the stem (the example defines only the type of a stem rather than the stressing paradigm). For more detailed definitions of the type of stem see (Kasparaitis, 2000).

Only those combinations of the type of a stem and the stressing paradigm are of significance in formulating the rules, which occur often enough. Only the combinations

Table 3

Compatibility of rules that have stressing paradigms 1 and 3 with those that have stressing paradigms 2 and 4

namas	brolis	arklys	ranka	vyšnia	bitė	krosnis	žvėris	sodžius	geras	žalias	didelis	medinis	gražus	Example defining the type of stem	
1 3	**1**	**3**	**1 3**	**1 3**	**1 3**	**1 3**	**3**	**1**	**1 3**	**1 3**	**3**	**1**	**3**	**str. paradigm**	
2 2 2 2	0 0	0 0	2 2 2 2	0 0 0 0	2 2 2 2	2 2 2 0	2 0	0 0	2 2 2 2	0 0 0 0	2 2	2 2	2 2	**2** **4**	namas
0 0	2	2	0 0	2 2	2 2	2 2	2	2	2 2	2 2	2	2	2	**2**	brolis
0 0	2	2	0 0	2 2	2 0	2 2	2	2	0 0	2 2	2	2	2	**4**	arklys
2 2 2 2	0 0	0 0	2 2 2 2	0 0 0 0	0 0 0 0	2 2 2 0	2 0	0 0	2 2 2 2	2 2 2 2	0 0	0 0	2 2	**2** **4**	ranka
0 0 0 0	2 2	2 2	0 0 0 0	2 2 2 2	2 2 2 0	2 2 2 2	2 0	2 2	0 0 0 0	2 2 2 2	2 2	2 2	2 2	**2** **4**	vyšnia
2 2 2 2	2 2	2 0	0 0 0 0	2 2 2 0	2 2 2 2	2 2 2 0	2 0	2 2	0 0 0 0	2 2 2 0	2 2	2 2	2 0	**2** **4**	bitė
2 0	2	2	2 0	2 2	2 0	2 2	2	2	2 0	2 2	2	2	2	**4**	krosnis
2 0	2	2	2 0	2 0	2 0	2 2	2	2	2 0	2 0	2	2	2	**4**	žvėris
2 2 2 2	0 0	0 0	2 2 2 2	0 0 0 0	0 0 0 0	2 2 2 2	2 2	0 0	2 2 2 2	0 0 0 0	0 0	0 0	2 2	**2** **4**	sūnus
0 0	2	2	0 0	2 2	2 2	2 2	2	2	0 0	2 2	2	2	2	**2**	sodžius
2 2	2	0	2 2	0 0	0 0	2 0	0	0	2 2	2 2	0	2	2	**4**	geras
0 0	2	2	2 2	2 2	2 0	2 2	0	2	2 2	2 2	2	2	2	**4**	žalias
2 2	2	2	0 0	2 2	2 2	2 2	2	2	2 0	2 2	2	2	2	**4**	didelis
2 2	2	2	0 0	2 2	2 2	2 2	2	2	2 2	2 2	2	2	2	**2**	medinis
2 2	2	2	2 2	2 2	2 0	2 2	2	2	2 2	2 2	2	2	2	**4**	gražus

that have at least 5 realisations in our dictionary (over 53000 words) are presented in the Tables.

The figures have the following meanings in the Tables: 0 means that rules are compatible irrespective of the stress position and the accent, 1 – that rules are compatible if they have the same stress position and accent, 2 – that rules are non-compatible.

The results are presented for those types of the stem only, which have the stems that do not end in the letter "j". If the stem ends in the letter "j" in the rule, the latter is compatible with all the rules having stems ending in other letters. The stems that end in the letter "j" have been analysed separately and three analogous tables have been obtained. They are not presented here, however, they will be used in this paper. Those Tables contain fewer compatible rules.

By means of those Tables it is easy to check whether two rules are compatible or not. E.g., Table 1 clearly shows that rules "*an-A_1" ("lie∼kana") and "*an-A_3"

(“dovana`”), are not compatible (cell “ranka 1” – “ranka 3”). Table 2 indicates that rules “*ai∼n-Ė_2” (“mišrai∼nė”) and “*ai∼n-IS_2” (“riestai∼nis”) are compatible because they have the same stress position and accent (cell “bitė 2” – “brolis 2”). Table 3 proves that rules “*i′en-A_1” (“avi′ena”) and “*ie∼n-Ė_2” (“uogie∼nė”) are compatible (cell “ranka 1” – “bitė 2”).

6. Correctness of Sets of Rules

DEFINITION. The set of rules enabling nouns and adjectives to be stressed with the same accuracy as in the cases when the whole stems are used is called a correct set. If the opposite case the set is called an incorrect set. The rules that make the set incorrect are called incorrect rules.

Only correct sets of rules will be investigated.

In the paper (Kasparaitis, 2000) it has been indicated that the following result could be obtained by stressing a word:

1) the word is stressed correctly;

2) the word is not stressed because it was not found in the dictionary;

3) the word is not stressed because there are many stressing variants;

4) the word is stressed incorrectly.

To verify the correctness of the set of rules, first and foremost it should be made clear what results could be obtained when the set of the whole stems is used to stress the same set. Let us take all the stems (one by one) and find all the matching stems. It is obvious that at least one stem with the same type of the stem, the stressing paradigm, the stress position and the accent will be found, i.e., we will find the stem itself. This means that results 2 and 4 are impossible. Besides, other matching stems can also be found. They can be either compatible or not with the stem being searched for. If the stem that is searched for is compatible with all the stems found (or no more stems have been found), the word can be stressed in one way only, if it is incompatible – there are many ways to stress it.

The following algorithm was used to verify the correctness of a set of rules: let us take one stem from the set of the whole stems and find all the matching rules in the set of rules. At least one rule with the same type of the stem, the stressing paradigm, the stress position and the accent must be found. Besides, other rules may also be found. They can be either compatible or not with the stem being searched for. Thus, the word might be stressed either in one way only or in many ways. The same results (stressing the word in one way or in many ways) must be obtained with the search for this word made in the dictionary of the set of the whole stems. In comparing these results it is only the fact compatibility or incompatibility that is of importance. Neither the number of rules found nor the number of compatible or incompatible rules is of any significance. The set of rules is correct provided that the results are the same for all the stems.

In case the set contains an incorrect rule, the latter can be eliminated from the set and all the whole stems that match this rule and that have the same type of the stem, the

stressing paradigm, the stress position and the accent can be entered in its place. Having replaced all incorrect rules in this, any set of rules can be made correct.

All sets of rules in this paper have been proved to be correct.

7. Formulation of Rules Manually

This chapter explains what rules have been formulated manually and why they can not be formulated automatically. A comparatively large dictionary of stems of nouns and adjectives containing words that are stressed in the same way already exists and they can be used to formulate stressing rules.

There are some infixes in the Lithuanian language, which are used to build new words that are stressed by means of the same stressing paradigm. However, the stress falls on the primary word (a string of letters) rather than on the infix (the end of the stem), with the stress position and the accent being retained. It makes no sense to create the algorithm for the automatic formulation of the rules realising these infixes because:

1) it is very difficult to recognise these infixes automatically and to define if words are stressed in the same way. This is because the string of letters defines the stress position and the accent, so the stress position and the accent can differ, e.g., "pu`sininkiškas" and "že∼miškas";

2) the string of letters must satisfy very strict restrictions because it must define the stress position and the accent. The restrictions that the string of letters should satisfy must be detected automatically. In other words, it must be detected automatically what string of letters this infix can be added to;

3) in case the attempt is made to formulate the rules automatically without taking into account the fact that these words retain the stress position and the accent of the primary word, the result is too many rules that reflect no principles of word-building at all;

4) such infixes are not abundant.

It was decided to formulate a certain set of rules manually. With this end in view principles of word-building by means of infixes presented in grammars of the Lithuanian language (Ambrazas *et al.*, 1996; Ulvydas *et al.*, 1996; Vaitkevičiūtė, 1997) have been analysed (with the exception of word-building of international words) and a list of restrictions or parameters applied to the string of letters has been compiled (the name of the parameter and its possible meanings):

"Where to search for". DB – to search in the database of stems of nouns and adjectives; VE – to search in the database of stems of verbs of present tense; VK – to search in the database of stems of verbs of the past tense; VB – to search in the database of stems of verbs of the infinitive; N – to search in no databases.

"Groups of types of stems". B – any adjective; BJ – an adjective that ends in the letter 'j'; BK – an adjective that does not end in the letter 'j'; D – any noun, DV – a noun of the masculine gender, DM – a noun of the feminine gender. If no meaning is given – to search among all types of stems.

"Stressing paradigm". K1 – stressing paradigm 1 or 2; K3 – stressing paradigm 3 or 4. In case no meaning is indicated – any stressing paradigm.

"Rejection of infix". IA – infixes must been rejected. This parameter defines if the infixes "ij" and "ik" must be rejected before the new infix has been added, e.g., "policij-a"+"-inink-as"="policinink-as", "matematik-a"+"-išk-as"="matematišk-as". In case no meaning is indicated – infixes are not to be rejected.

"Number of syllables". S1 – the stem is monosyllabic; SD – the stem is polysyllabic. If no meaning is indicated – the stem can have any number of syllables. This parameter can be used both together with nouns and adjectives and verbs.

"Prefix". P+ – a verb must have a prefix; P- – a verb can have no prefix. In case no meaning is indicated – a verb can either have a prefix or not.

"Class of conjugation". AOĖ – a verb belongs to the class of conjugation "o-ė" (the third person of the present tense has the ending "-o" and the third person of the past tense has the ending – "-ė"); AN – a verb does not belong to the class of conjugation "o-ė". If no meaning is indicated – any class of conjugation.

Then infixes were selected from the above mentioned grammars taking into account the following criteria:

1) the stress of the word with an infix does not fall on the infix and it is stressed according to stressing paradigm 1 or 3;

2) it is easy to define what restrictions are satisfied by the remaining part of the stem (a string of letters). In other words, what string this infix can be added to. E.g., the infix "išk" ("vaik-išk-as", "moter-išk-as") can be added to the stem of the noun and retain its place of the stress and the accent, however, it is not clear what string the infix "an" ("darg-an-a", "sam-an-a") can be added to;

3) once the rule has been formulated for an infix, the dictionary is to contain quite a number of words (over 10) that are stressed according to that rule, i.e., the words that match the ending of the stem, satisfy the restrictions for a string of letters, have the same type of the stem, the stressing paradigm, the stress position and the accent;

4) the dictionary should contain as few as possible words that are stressed erroneously according to that rule, i.e., the words that match the ending of the stem, satisfy restrictions on a string of letters, however, which are incompatible with the rule. Prior to checking the compatibility, the stress position and the accent of the rule is replaced with the stress position and the accent defined by a string of letters because the stress position and the accent are not defined in these rules.

The number of rules that have been formulated manually total 16. They are presented in Table 4.

Having rewritten the rules presented in the Table 4, they will look as follows:

"IA K1 DB*inink-AS_1" ("poli`cininkas"),

"AN SD VK*im-AS_1" ("kanki`nimas"),

"P+ VK*ėl-IS_1" ("pabė′gėlis").

The above listed restrictions can be successfully used in other rules in which the stress falls on the ending of the stem. E.g.:

"K3 DB*inin∼k-AS_2" ("darbinin∼kas"),

"S1 DV DB*e∼l-IS_2" ("name∼lis"),

"SD DV DB*ė∼l-IS_2" ("katinė∼lis"),

Table 4
Manually formulated rules

Rejection of infix	Stres.paradigm	Group of types of stem	Where to search for	Ending of stem	Type of stem	Stres.paradigm	Number of stressed words	Example
IA	–	–	DB	išk	AS (adj.)	1	616	vaikiškas
IA	K1	–	DB	inink	AS (noun)	1	375	mokslininkas
IA	K1	–	DB	inink	Ė	1	343	mokslininkė
IA	K1	–	DB	in	IS (adj.)	1	222	metinis
–	–	B	DB	ul	YS	3	20	gyvulys
Class of conjugation	Prefix	Number of syllables	Where to search for	Ending of stem	Type of stem	Stres.paradigm	Number of stressed words	Example
–	P-	S1	VE	ul	YS	3	64	snaudulys
–	P-	S1	VE	es	YS	3	64	judesys
AN	–	SD	VK	im	AS (noun)	1	3987	bauginimas
AOĖ	–	SD	VK	ym	AS (adj.)	1	850	rašymas
–	P+	–	VK	ėl	IS (noun)	1	272	numirėlis
–	P+	–	VK	ėl	Ė	1	262	numirėlė
–	–	S1	VK	in	YS	3	151	rašinys
–	–	SD	VB	toj	AS (noun)	1	524	mokytojas
–	–	SD	VB	toj	A	1	514	mokytoja
–	–	SD	VB	tin	IS (adj.)	1	47	statytinis
–	P-	–	VB	sen	A	1	36	rašysena

“AN S1 VK*i`m-AS_2” (“neši`mas”),
“IA DB*i′ng-AS_1” (“darbi′ngas”),
“BK DB*ia′us-IAS_1” (“geria′usias”),
“BJ DB*a′us-IAS_1” (“gaja′usias”),
“B DB*e`sn-IS_4” (“gere`snis”).

There is one important difference between the rules in which the stress falls on the ending of the stem and the rules with the stress falling on the string of letters: when checking restrictions on a string of letters, in the first case it is suffice to find one matching string, whereas in the second case – all matching strings must be found and it must be checked whether they all have the same stress position and accent.

8. Automatic Formulation of Rules

First and foremost, the idea of how it is possible to formulate stressing rules automatically and how to minimise the number of rules (together with the remaining stems) should be explained. Let us analyse the following group of words: "matado`ras", "semafo`ras", "cho`ras", "vo′ras", "no′ras", "bajo∼ras". The last two letters of the stem of all these words coincide (4 last letters in the word) and the third letter from the end unambiguously defines stressing of the word (stressing paradigms are not taken into account). Consequently, the rules with 3 letters in the stem rather than the whole stems can be used to stress these words: "*do`ras", "*fo`ras", "*ho`ras", "*vo′ras", "*no′ras", "*jo∼ras". Let us group these rules in such a way that the rules defining the same stressing should belong to one group. If we replace the largest group by one short rule (in this case "*do`ras", "*fo`ras", "*ho`ras" is replaced by "*o`ras") and leave other groups unchanged, it is obvious that the smallest set of rules will be obtained.

Now let us assume that we have a large set of rules with the length of the ending of the stem being equal to M. These rules can be joined in groups (in the way analogous to that described above) in which the $M-1$ letter in the ending of stem is the same. Then each group is divided into subgroups with the same stressing, the largest subgroup is found and one rule of the length $M-1$ for this subgroup is formulated. However, this algorithm can be repeated once again for all these rules of the length $M-1$ and so on. Every iteration produces the minimal number of rules.

Now a complete algorithm of formulating the rules by means of a large dictionary can be presented:

1) All the words that cannot be used for the formulation of rules must be entered on the list of exceptions. These are the words that have non compatible stems (the explanation why will be presented later).

2) All the words with the stress position farther from the end of the stem than M (where M is the length of the rule) must be entered on the list of exceptions. We choose this value at the beginning of the process. The value must be big enough to make all the words with the same ending of the stem compatible. $M=6$ was used in all experiments. The larger M the smaller number of rules should be obtained, however, after a certain limit has been reached the number of rules no longer decreases and the number of calculations increases.

3) All the remaining words are to be sorted out according to the last M letters of the stem, the type of the stem, the stressing paradigm, the stress position and the accent.

4) The stems that have been sorted out are to be divided into groups with the stems having the same $M-1$ last letters.

Then two variants of this algorithm (A and B) have been analysed. This relates to the fact that words with different types of the stem can have the same endings, which causes certain confusion. Variant A is quite simple, however, it enables us to formulate only one rule per ending of the stem for all stem types. Variant B is much more complicated but it sometimes allows us to have one rule per type of the stem.

A

5) Divide the groups formed in step 4 into subgroups so that only the words with the same type of the stem, the stressing paradigm, the stress position and the accent should belong to one subgroup.

6) Find the largest subgroup in each group and formulate one rule for this subgroup.

B

5) Divide the groups of words into smaller groups with the same type of the stem. Divide these smaller groups into subgroups so that only the words with the same stressing paradigm, the stress position and the accent should belong to one subgroup.

6) Find the largest subgroup within each small group and select those subgroups from the large group that produce all compatible rules and the sum of all elements in these subgroups is the largest. The generation of all possible combinations of the subgroups and the search were used for this purpose. The minimal number of rules would be obtained if all subgroups were used in the search (not only the largest ones). This was not done to speed up the calculation.

Formulate one rule per subgroup for all the groups that were selected within each group.

7) Newly formulated rules must be left for future processing. Other rules that were not used in formulating the new rules should be entered on the list of exceptions.

8) The length of rules should be reduced.

9) Steps 3–8 must be repeated as long as the length of rules M decreases to the desired length. The iterations were repeated until M became equal to 2.

10) (Optionally) All the rules that are used for stressing only one word must be found and changed back into a full stem.

The list of exceptions and the rules that were left in the last iteration form the final set of rules.

Table 5 shows how this algorithm works when all the words have the same type of the stem and $M = 5$. In this example no account is taken of the fact whether the stress falls on the ending of the stem.

Table 5

Automatic formulation of rules

Words	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Results
paklo∼tė	*a_klo∼tė	*k_lo∼tė			*klo∼tė
įklo∼tė	*į_klo∼tė				
poliglo`tė	*i_glo`tė	*g_lo`tė	*l_o`tė		*lo`tė
pilo`tė	*p_ilo`tė	*i_lo`tė			
despo`tė	*e_spo`tė	*s_po`tė	*p_o`tė		*po`tė
mazgo∼tė	*a_zgo∼tė	*z_go∼tė	*g_o∼tė	*o∼tė	*o∼tė
garbano∼tė	*b_ano∼tė	*a_no∼tė	*n_o∼tė		
bukapro∼tė	*a_pro∼tė	*p_ro∼tė	*r_o∼tė		
bepro∼tė	*e_pro∼tė				
pu`sprotė	*`s_protė				*`sprotė

It should be noted that sometimes it is impossible to find the largest subgroup because some subgroups have the same number of elements. This situation occurs more often if the subgroups are small, e.g., if the group consists of two differently stressed words (one subgroup = one word). If we want to have a minimal number of rules we should keep in mind such situations and in the next iteration verify what results are obtained by making use of the first subgroup for formulating a rule and what results are obtained by means of the second one. Seeking to simplify the algorithm this was not done and one subgroup was chosen.

If the set of rules has been build using the algorithm A, the search in the set of rules is simpler. Having found one matching rule (the rule that matches the ending of the stem) of the length equal to M, only those rules must be sought for, that are not shorter than M. Therefore it is possible to line the rules in the decreasing sequence and terminate the search after the first shorter rule has been found.

If method B is used, all the matching rules (the rules that match the ending of the stem) are to be found irrespective of their length. Then the rules that match the ending must be found and the longest rule (or several longest rules) must be selected out of them. It is impossible to select the longest rules at once because there might be compatible words that have only different endings. E.g., the rule "*ju′ost-A_1" was formulated by using the words "fotoju′osta", "perfoju′osta", "videoju′osta". Another rule "*uo∼st-Ė_2" was formulated by using the words "pašluo∼stė", "prijuo∼stė", "šluo∼stė". If we looked for the longest rule only in order to stress the word "prijuo∼stė", we would find the rule "*ju′ost-A_1". However, later, when looking for the ending defined by this type of the stem, we would find no matching ending and the word would remain unstressed.

Due to this reason another interesting situation can occur when a word is stressed correctly by applying the rule that was formulated for stressing the word having another type of the stem. E.g., the rule "*le`t-IS_2" has been formulated by using the words "lengvaatle`tis", "sunkiaatle`tis". Another shorter rule "*e`t-Ė_2" has been formulated by using the words "brune`tė", "rake`tė", "table`tė". If we tried to stress the word "table`čių" we would find the rule "*le`t-IS_2" (because there exists the form "lengvaatle`čių") which defines both the stress position and the accent correctly. But this rule has been formulated by using the words with another type of the stem. This is because both rules define the same stress position and accent for the words with this ending.

9. Results of Experiments of Formulation of Rules

We started our experiments having a dictionary containing 53149 words (called Dictionary Z).

Words that have the same stems and types of the stem, e.g., the said "pyli`m-as" and "py′lim-as", cannot be used for the formulation of rules. If one of such words were used (only one word can be used because only one rule can be formulated for one type of the stem) to formulate the rule, e.g., "*i`m-AS_2", the rule "py′lim-AS_1" corresponding to the second word would be used in all cases because it is longer (a longer rule is given a

higher priority). Thus, many stressing variants would not be recognised and words would always be stressed in a single way (sometimes erroneously). If 64 words with the same stem and type of the stem were removed from the Dictionary Z, the latter would contain the remaining 53085 words. However, this is not enough since some words with the same stems and different types of the stem can not be used to formulate the rules. Having removed all 12984 words with the same stems 40165 words would remain. However, some of these words can be used for the formulation of the rules. It is suffice to remove non-compatible words with the same stems. The number of such words removed totalled 3488 (Dictionary N) and 49661 words (Dictionary L1) were left. The latter words will be used in future experiments.

16 rules were formulated manually (Dictionary R). The words that can be stressed by means of manually formulated rules, i.e., the words that match the ending of the stem, satisfy the restrictions on a string of letters, have the same type of the stem, the stress paradigm, the stress position and the accent, can be removed from the dictionary. Dictionary L1 contains 8283 such words (Dictionary I). Moreover, the words that are stressed erroneously by applying manually build rules, i.e., the words that match the ending of the stem, satisfy the restrictions on a string of letters but are not compatible with the rule, cannot be used for the formulation of the rules. Prior to checking the compatibility, the stress position and the accent in the rule must be replaced with the stress position and the accent defined by a string of letters. 582 such words (Dictionary K) were found in Dictionary L1. If these two groups of words were removed, 40796 words would remain (Dictionary L2), which would be used for an automatic formulation of the rules.

Besides, the words, that match the ending of the stem, have the same type of the stem and the stressing paradigm, but which fail to satisfy the restrictions on a string of letters are of interest when checking the restrictions on manually build rules. The number of such words amounts to 2429 (Dictionary A). If the number of words in each dictionary were designated by means of corresponding lowercase letters the ratio $(r + a)/i$ could be used to measure the quality of manually build rules.

By means of algorithm A, two experiments were carried out with Dictionary L2 in which the rules have been formulated automatically, one rule per ending of the stem (for all types of stems). If several groups contained the same number of stems, the first group was used in one experiment and the last group – in the second one. The number of stems and rules obtained was 16825 and 2438 respectively (the total of 19263 records) in the first experiment and 16546 stems and 2308 rules (the total of 18854 records) in the second one.

By means of algorithm B two another experiments were carried out with Dictionary L2 in which rules were formulated automatically, one rule per ending of the stem per type of the stem rejecting incompatible rules. If several groups contained the same number of stems the first group was used in one experiment and the last group – in the second one. The total of 12338 stems (Dictionary M) and 2475 rules (Dictionary T) was obtained (14813 records in total) in the first experiment and 12428 stems and 2517 rules (14945 records in total) in the second one. The results of the first experiment will be treated as final results of an automatic formulation of a set of rules.

Instead of Dictionary Z a joint dictionary containing Dictionaries N, K, M, R and T can be used for the automatic text stressing. The size of the joint dictionary is $3488 + 582 + 12338 + 16 + 2475 = 18899$. Consequently, the size of the initial dictionary (Z) has been reduced by 64,4%.

10. Automatically Formulated Rules

Having looked through the automatically formulated rules, one can notice that the rules have been formulated for the words that have:

1) the same infixes of the Lithuanian language,
2) the same infixes of international words,
3) the same second component of compound words,
4) the same second component and linking vowels of compound words,
5) simply the same endings of the stem.

Two examples of each case are presented in Table 6. Besides, the Table shows how many words are stressed by applying each rule.

It would be complicated to formulate so many different rules manually with the help of grammars because grammars usually present rules for stressing words with the same infixes, whereas in this paper rules for stressing compound words and words that are stressed in the same way have been formulated. Besides, rules that are used to stress only one or very few words are presented in grammars. E.g., (Ulvydas *et al.*, 1996) wrote that adjectives with the infix "-yn-as" are stressed using the first stressing paradigm. However, we have found only one such adjective "mėlynas" in the dictionary available, so it is more

Table 6
Automatically formulated rules

No.	Rule	Examples	Number of words
1	"*e∼l-IS_2"	"rage∼lis", "svirbe∼lis"	106
	"*nin∼k-AS_2"	"darbinin∼kas", "kaklinin∼kas"	93
2	"*i`zm-AS_2"	"komuni`zmas", "sociali`zmas"	604
	"*a∼cij-A_1"	"dota∼cija", "reputa∼cija"	697
3	"*e`tr-AS_2"	"termome`tras", "milime`tras"	170
	"*en∼tr-AS_2"	"epicen∼tras", "metacen∼tras"	8
4	"*ė′tyr-A_1"	"pelkė′tyra", "upė′tyra"	6
	"*o′svaid-IS_1"	"kulko′svaidis", "mino′svaidis"	4
5	"*i`ž-IUS_2"	"avi`žius", "smali`žius"	3
	"*u`š-AS_2"	"du`šas", "tu`šas"	3

convenient to treat such cases as exceptions rather than rules. Another problem is that grammars do not usually give the number of words stressed according to a certain rule (there are exceptions, e.g., (Pakerys, 1991)). If words with a certain infix can be stressed in many ways, it is difficult to decide which rule is more important to stress as many words as possible. E.g., (Vaitkevičiūtė, 1997) indicates that the infix "-on-as" of the noun can be stressed in three ways. 298 cases of stressing "-o∼n-as" ("baro∼nas"), 45 – "-o`n-as" ("vago`nas") and only 4 cases of stressing "-o′n-as" ("lavo′nas") have been found in the dictionary available.

No such problems are encountered when the automatic formulation of rules is used.

11. Experiments of Text Stressing

Before using a set of rules to stress nouns and adjectives, certain priorities should be assigned to the rules. As the whole stems present in the set of rules usually form exceptions, they will be assigned the highest priority (let us denote it N). Lower priority will be assigned to the manually formulated rules. Assume the length of the rules (the number of letters in the ending of the stem) k to be less than K. Then the manually build rules will be assigned the priority $M + k$ where $M + K < N$. The lowest priority (k) will be assigned to the automatically build rules. Consequently, $M > K$.

If we know in advance that the word to be stressed is a noun or an adjective, the following stressing algorithm can be applied:

1) find all the rules that match the word we want to stress;

2) select the rules that have the highest priority;

3) verify if the word is stressed in the same way using all these rules.

If we stress any text, stressing of verbs and non-inflectional words should be added to the stressing algorithm of nouns and adjectives described above. For this purpose let us assume the verbs, non-inflectional words, the whole stems of nouns and adjectives as well as manually formulated rules to have an equally high priority and the automatically formulated rules – to have a lower priority. I.e., if a certain word matches the stressing rule of a verb (or a non-inflectional word) and the automatically formulated rule, the automatically formulated rule will be assumed to match accidentally and it will not be taken into account.

Seeking to establish the reliability of the algorithms presented in this paper, experiments were carried out with the texts covering about two pages of social and political journalism and fiction. The same texts were used in the paper (Kasparaitis, 2000). Results are presented in Table 7. Results from the said paper are presented in the upper part of the Table for comparison.

As can be seen, the number of correctly stressed words has been increased and the number of words that were not found in the database has been decreased because the diminutives (e.g., "gėlelė") and some surnames (e.g., "Dudėnas") have been stressed. On the other hand, the number of incorrectly stressed words has been increased and some words with a number of stressing variants have not been recognised. This is because some surnames have been stressed erroneously (e.g., "Deguti′enė" was stressed

Table 7

Results of text stressing experiments

		Stressed correctly	Stressed incorrectly	Unstressed. Not found in the database	Unstressed. Many stressing variants	The total
Whole stems used						
Social and political journalism	Number of words	341	0	15	57	413
	%	82,57%	0%	3,67%	13,80%	100%
Fiction	Number of words	406	1	6	85	498
	%	81,53%	0,20%	1,20%	17,07%	100%
Stressing rules used						
Social and political journalism	Number of words	343	8	8	54	413
	%	83,05%	1,94%	1,94%	13,08%	100%
Fiction	Number of words	412	2	2	82	498
	%	82,73%	0,40%	0,40%	16,47%	100%

"Degutie∼nė"). Besides, the automatically formulated rules must sometimes have the same priority as verbs and non-inflectional words. E.g., the word "visuomenės" has been stressed as the future tense "visuomenė∼s" of the verb "visuomenėti", however, according to the automatically formulated rule "*u′omen-ĖS" it must also be stressed "visu′omenės".

These errors are possible to be corrected in the following way:

1) a certain set of names, surnames and names of places must be included in the dictionary before building the set of rules.

2) restrictions on some automatically build rules must be made stricter in the manual way. These rules will be transferred in this way to the class of manually formulated rules. Another method is to enter the whole stem in the set of rules. However, the problem arises how to recognise such automatically formulated rules. Probably the surest method is to generate all possible grammatical forms of verbs (adding non-inflectional words to them) and to stress them using dictionaries of verbs, non-inflectional words, nouns and adjectives. Then to stress them using automatically formulated rules and find out in what cases the results differ.

12. Results, Conclusions and Trends of Future Work

In conclusion, a summary of what has been done in the present work is made:

1) the form of stressing rules of nouns and adjectives has been defined;
2) a new concept that is of paramount importance to the formulation of the rules has been introduced, namely – the compatibility of rules;
3) a set of stressing rules has been created manually;
4) the algorithm for the automatic formulation of stressing rules has been created;
5) the volume of a dictionary of stems of nouns and adjectives that was used for stressing purposes has been reduced by 64,4%;
6) the results obtained by means of the stressing rules are similar to those obtained by applying the whole stems (the number of correctly stressed words has increased by approximately 0,9%, the number of incorrectly stressed words has increased by 1,0% and the number of unstressed words has decreased accordingly).

Conclusion. The algorithm for the automatic formulation of the stressing rules of nouns and adjectives can be successfully used to generate rules, and the rules created by means of this algorithm can be used to stress the text automatically. However, the interaction of these rules with the stressing rules of verbs and non-inflectional words is to be improved.

In the opinion of the present author, the following trends of the future research can be promising:

1) to create a similar algorithm for stressing of verbs;
2) to create a similar algorithm to stress words according to their beginnings (prefixes);
3) to generate all possible grammatical forms of nouns, adjectives, verbs, to add non-inflectional words to them and to automatically formulate their stressing rules;
4) to analyse algorithms that make it possible to select one stressing variant.

References

Ambrazas, V., K. Garšva, A. Girdenis *et al.* (1996). *A Grammar of Modern Lithuanian*. 2nd ed. Mokslo ir enciklopedijų leidykla, Vilnius (in Lithuanian).

Kasparaitis, P. (2000). Automatic stressing of the Lithuanian text on the basis of a dictionary. *Informatica*, **1**(11), 19–40.

Pakerys, A. (1991). *Stressing of International Words.* Šviesa, Kaunas (in Lithuanian).

Ulvydas, K., *et al.* (1965). *A Grammar of the Lithuanian Language*. T.1., Mintis, Vilnius (in Lithuanian).

Vaitkevičiūtė, V. (1997). *Stressing of Common Lithuanian Language.* Šviesa, Kaunas (in Lithuanian).

P. Kasparaitis was born in 1967. In 1991 he graduated from Vilnius University (Faculty of Mathematics). In 1996 he has been admitted as a PhD student in Vilnius University. Current research interests include text-to-speech synthesis and image processing.

Lietuvių kalbos daiktavardžių ir būdvardžių automatinis kirčiavimas naudojantis taisyklėmis

Pijus KASPARAITIS

Šiame darbe nagrinėjamas lietuvių kalbos teksto automatinis kirčiavimas. Ankstesniame darbe autorius yra pateikęs lietuvių kalbos teksto automatinio kirčiavimo remiantis žodynu algoritmą. Šio darbo tikslas – patobulinti minėtą algoritmą įtraukiant formalias daiktavardžių ir būdvardžių kirčiavimo taisykles. Jomis naudojantis sukirčiuojami ir žodyne nesantys žodžiai: deminutyvai, pavardės, būdvardžių laipsniai. Darbe išanalizuota, kada patogiau taisykles sudaryti rankiniu būdu, o kada generuoti automatiškai. Jame aprašytas rankinis taisyklių sudarymo būdas bei pateiktas tokių taisyklių rinkinys. Be to, pateiktas automatinio formalių taisyklių generavimo iš daiktavardžių ir būdvardžių žodyno algoritmas.