Ľubomír Ď u r o v i č

FORMALIZING THE S/C FLEXION FOR THE AUTOMATIC MORPHOLOGICAL ANALYSIS

The computerizing of the JUBA-project was intended from the beginning as a means of handling a vast corpus containing many hundreds of thousands of words, partly phonetically and morphologically corrupt. Those corruptions differ from one informant to another, so no certain predictions can be made based on otherwise statistically sufficient parts of the corpus: all phenomena investigated must be physically available. That presupposes a) a normalizing of all texts to a convenient norm, enabling the computer search, b) a system of programs for searching in the corpus, based on the automatic morphological analysis and hence on the grammatical identification of the text words (i e text segments between two blanks).

1.0. The normalizing of the text was described in the paper by Ingemar Dahlstrand (Ingemar Dahlstrand 2.3.). By introducing the two signs - * and & - t h r e e levels of representation were introduced in the corpus: a) the first one, corresponding to the physically present text on the tape (of course, the phonetical deviations are not precisely transcribed, and cannot be precisely transcribed, by means of e g API signs, since the children's sounds are more or less individual within a certain "zone"; a paper on this phenomenon is under preparation) and they do not represent any known social standard; b) the second form, following the asterisk *, "the asterisk forms", where the words are restored as if they consisted of and only of S/C phonemes, written in the codified orthography (to this level belongs, of course, even "ě" representing all reflexes of the historical "jat"); c) the morphologically restored text following the "&", "the gamma forms". The automatic analysis and identification work on the last two representations.

The "asterisk representation" can be read by the machine, but not always analysed and, consequently, identified. The "gamma representation" is the primary object for computer operations, since it is normalized according to the rules of the standard S/C.

It is possible to work with those three representations in any

Slavica Lundensia 9, 1983, pp. 123-149

combination and to search for a combination desired, e g only those asterisk forms, where a "p" corresponds to the physically present (and transcribed) "ph" etc.

2.0. The automatic computer identification (ACI) of the text words yields their grammatical characteristics (analysis class, i e approximate part of speech; exhaustive paradigmatic information for the inflected parts of speech), their lemmatization and certain grammatical characteristics of the lemma (genus, animation for nouns; aspect and reflectivity for the verbs).

The ACI could be based on a dictionary of all paradigmatic word forms; but, since e g a verbal paradigm contains some 150 forms, such a dictionary would get enormous dimensions already with some 50 - 100 verbs. Therefore, an automatic morphological analysis (ACA) has been developed by JUBA for the open inflected parts of speech.

The ACI is done directly by the word form dictionary (OFLX); ACA is done by comparing the inflectional theme of a text word with the theme (or stem) dictionary (STLX) and each identity found by this search is compared with the dictionary of endings (ENLX), testing then the compatibility of the segments found: the grammatical identification is the result of this comparison.

The choice between the immediate ACI based on the OFLX and the complex ACA is a question of economy. In the JUBA project, we decided to use ACA for the three great open formal classes - the nouns (class 1), the adjectivals (= all congruent words such as adjectives, ordinal numerals, congruent pronouns: class 2) and the normally conjugated verbs (class 7). Verbs with isolated flexion, such as (je)sam, jesi ..., nisam, nisi ..., as well as personal pronouns, the interrogative pronouns (t)ko and što/šta, the potentially inflected numerals 2 - 4 etc are not computer analysed: all their forms are in the OFLX and will be directly identified.

3.0. The ACA of the substantives, adjectivals and verbs required an explicit, nonambiguous, computer adapted description of those word classes in the S/C. This task has never been undertaken earlier and therefore I now want to deal with some aspects of this work. I do not want to describe the "trial and error" stages preceding each decision - we had no clear idea about the optimal presentation of the linguistic data for the machine to begin with, and tried different schemes.

3.1.1. For a linguist one question must be clear from the very

beginning: the entities stated in a language by the linguistic analysis are suitable for processing only after having been adapted for the computer, i e formalized.

The only machine readable substance at present is the graphics (printed letters) and therefore S/C graphics in their actual orthography became the object of formalization. That means: no underlying entities can exist, there is no difference between automatic neutralizations and morphophonemic alternations as long as they are not distinguished in the S/C orthography. E g a word such as redak must have three theme allomorphs in the STLX (if the vocative is retku), i e r e d a k for combinations with the endings of N sg redak and G pl redaka, r e t k for G sg retka, D sg retku... etc and pl retkovi ... (se 3.2.) and r e c for N pl reci and D-L-I pl recima. It does not matter that some of the sound/letter changes in the theme are dŭe to the automatic neutralizations (/d/ -> [t] , [c]) and degemination ([cc] -> [c]), but some others to the theme alternations (/red#k-/-red#c-/: it is needless to emphasize that there does not exist any palatalization in the S/C).

3.1.2. The linguistic requirement of consistency in defining the items is not necessary for the machine. Let us exemplify it on the present tense of the verbs. The S/C has three main conjugations -e-, -i- and -a- type (e g nesem, učim, čitam). Linguistically, all "present stem" forms are segmented into three morphemes: stem, grammatical suffix (class index) -e-, -i- or -a- and the endings as -m, -š ..., -ći (gerund), -i (imper) etc. The generation of the forms, where this tripartition is not evident, is a special question, abundantly discussed in the last decades after the famous article "Russian Conjugation" by R. Jakobson in Word 4, 1948, 155 ff. There is no doubt that the same endings -m, -š etc appear in all three conjugations.

For the computer formalization the question of the segmantation is solved in another way: the only criterion is to diminish the quantity of items.

In this respect the three conjugations differ from each other, viz:

piš em	nos im	vs	čita m
piš eš	nos iš		čita š

piš e	nos i	čita ∅
piš emo	nos imo	čita mo
piš ete	nos ite	čita te
piš u	nos e	čita ju

Thus, in the a-conjugation, we obtain the present stem, identical to the infinitive stem, i e all forms in this conjugation can be found by a single stem allomorph in STLX-7 and any second allomorph, such as nos-/nosi-, pisa-/piš- is not required. Having in mind the vast productivity of this conjugation, this is an important saving of e g memory space. The new ending series -m, -š, -∅ ... does not increase the number of items in the ENLX, since after the principal decision in favour of the bipartition (instead of a possible tripartition, see 3.2) we had the choice between -m, -š, -∅ ... and -am, -aš, -a, both of them different from -em, -eš, -e ... and -im, -iš, -i...

3.2. The central question for the conception of ACA is the question of how to segment the word forms, viz bipartition vs tripartition. The S/C verbal forms consist of three parts: the stem, the grammatical suffix (also called "class index") and the ending, the suffix and the ending being the grammatical portion of the word form. The S/C nominal forms have been, for about 1000 years, bipartite (the flexion theme and the ending), but there is a clear tendency in the S/C to generalize the use of grammatical suffixes (also called "grammatical redudant morphemes", "predesinential suffixes", Russ. "rasprostranenie osnovy"), such as -in (Švedjan-in-∅, Švedjan-in-a), -et (jar-∅-e, jar-et-a) in sg and -ov/-ev (brod-ov-i, oč-ev-i so called "long plural"), -ad (jar-ad-∅) in pl.

Because of this general tripartiteness (or tendency towards it) of S/C word forms it was quite natural to pose the question, whether the ACA search could not work on three subsequent segments. In other words: whether one of the dictionaries could contain "bare" morphemes such as -in-, -ov- etc, in order to enable the ACA to identify both brodi and brodovi from the same inventory of items.

This possibility was rejected because of the following arguments: a) the tripartition is extremely uneconomic for the computer since it means a two-dimensional search for stem-ending compatibility, viz stem vs grammatical suffix and then stem + grammatical suffix

vs ending; b) the grammatical suffixes in the singular can be conceived as a part of the theme (turčin-Ø) and of one of the theme allomorphs respectively (jare-/jaret- as redak-/retk-/rec-); c) since the "long plural" suffixes -ov and -ev combine with only one series of endings, namely N-V -i, G -a, D-L-I -ima, A -e and always with it, it turned out to be very advantageous to conceive those combinations as separate endings -ovi, -ova, -ovima, -ove, and -evi, -eva, -evima, -eve respectively, to place them all in the ENLX and to symbolize the whole of those series by a single sign xo and xe respectively in the STLX.

The bipartition of the word forms thus became a universal principle for building up all the lexicons for ACA. Each lexeme is represented in the respective dictionary by all its theme allomorphs and for each of them are noted the endings which combine with it (see the paper by Ingemar Dahlstrand, 2.1.). E g:

redak	1	mi	redak	ns			as				gsa		
retk	1	mi	redak		gsa	dsu		vsu	iso			ape	xo
rec	1	mi	redak							npi		dpi	

(For interpreting the signs, see 4.3.).

This is to be compared to our earlier tripartite conception:

	stem	gram suff	ending
redak	20	sg 1	3
		pl 1,2	1

where every number referred to an item in our earlier tentative lists (stem type list, grammatical suffix list, ending series list).

4.0. As we said above, the formalizing for the ACA requires full explicitness. It means that for each ending in each paradigm (i e for each grammatical value of a homonymous ending as well) its grammatical value must be explicitly stated.

Here the question arises how the stem dictionaries (STLX) can be built for such parts of speech as e g a verb, containing about 150 forms (aspects and passive forms not included). Ascribing explicitly each one of the 150 endings to one of the 1 - 4 theme/stem allomorphs for each verb is in all respects practically impossible or at least would be extremely uneconomic. How then can the para-

digmatic systems of S/C be exploited in order to simplify the data in the STLX, leading to the identification of each word form through the ENLX?

It turned out that the typological difference between the verbs and the adjectivals on the one hand (4.1.) and the nouns (4.2.) on the other, enabled/forced us to build the STLX:s for those two groups in two different ways.

4.1.1. Both the verbal and the adjectival paradigm consist of many subparadigms, i e present, future, aorist, imperfect, imperative, gerunds and infinitive; passive participle; subparadigms consisting of L-form (the S/C "pridjev radni") and auxiliaries as well as of infinitive and auxiliaries. Each synthetic tense subparadigm consists of six person forms, each analytic tense subparadigm of six person forms multiplied by three genders (bio sam, bila sam, bilo sam etc), each infinitive subparadigm of one single form etc. In the same way within the adjectivals each masculine, each feminine and each neuter subparadigm consists of seven case forms in the singular and seven in the plural, at least when the full ("determined", "određene") forms are concerned, etc.

Because of the internal development in S/C within each of those subparadigms
a) the flexional theme is free from consonant alternations: the only exception is the closed verbal class peći, i e with the historical velar in the stem;
b) all the endings are predictable from any one of them: e g a verbal -a (3rd sg pres) predicts without exception -am, -aš, -amo, -ate, -aju in the other five present tense forms, an adjectival -emu (dat sg m or n) predicts in the same way -eg or -ega (gen sg), -i (nom sg masc, nom pl masc), -im (instr sg), -ih (gen pl) etc: in the adjectival flexion even all subparadigms are predictable from each other.

Because of this, entire subparadigms can be symbolized by one identification number and the whole verbal paradigm (synthetic forms) is a combination of 10 such identification numbers (see Ingemar Dahlstrand, 2.1.2). The following subparadigms are symbolized by the first digits in the verbal identification numbers.
1 - infinitive
2 - L-form (the base form of preteritum)
3 - aorist
4 - present
5 - imperative
6 - present gerund

7 - past gerund
8 - passive participle
9 - future (synthetic, e g čitaću)
q - imperfect.

The second digits symbolize the 2 - 3 ending series in contemporary S/C, e g:
41 - em, eš, e, emo, ete, u
42 - im, iš, i, imo, ite, e
43 - m, š, Ø, mo, te, aju (see 3.1)

Concerning the rare cases, when a subparadigm ending series must be split between two stem allomorphs, see Ingemar Dahlstrand, 2.1.0.

4.1.2. At the present stage the analytic ("periphrastic") forms are not identified as such: the ACA identifies only L-forms (and not past čitao sam, conditional čitao bih, pluperfect bijah čitao, or future II budem čitao), only infinitive (and not future I čitati ću), only unspecified voice (and not active vs passive, viz radi vs radi se).

The decision whether the forms sam, si, je ... are auxiliaries belonging to the analytic past čitao s a m or copula forms ja s a m ovdje, must now be made manually in the interactive program SHOW. Automatic solutions do not, however, seem to be an insoluble problem. Since the past, future and conditional auxiliaries are enclitics, bound in different ways within the frame of a predication (a text segment between two + in our stored records), the identification of those analytical forms is made possible by search subroutines within a predication starting when an L-form or infinitive are identified.

A presupposition for identifying passive forms ("the reflexive passive") has been posed by reserving one character in the STLX-7 for reflexivity: "7 i0" vs "7 ir" means an imperfective (i) non reflexive (0) vs imperfective reflexive (r) verb (7). Here, "reflexive" signifies reflexivum tantum ("deponentium", e g smejati se) or reflexive verb (such as biti se vs biti nekoga), but not a reflexive f o r m (passive or impersonal, as Na Uskrs se u selu boje jaja). The automatic processing of those analytic forms is planned for the next stage of work.

4.1.3. In the adjectivals, the predictability of all the word forms from the theme allomorph is almost absolute both in the positive and in the comparative: only in unique cases can the use

of endings with -o- or -e- (viz -oga, -om(u) vs -ega, em(u)...) be a question, e g stari, staroga but širi, širega. The meaning of comparative is always expressed by the theme allomorph only. The only non-predictable features are the existence of the short forms ("neodređeni oblik") and of the zero-vowel ("nepostojano a"). Therefore in this analytical class the maximal simplification is possible in the STLX: in many cases the whole gender-number word form complex can be symbolized by only one symbol in the STLX-2, e g ke (= comparative, ending series -ega, -em(u),... -e, -ej, ...).

4.1.4. The identification numbers/signs used in the STLX are then specified in the ENLX as can be seen in the article by Ingemar Dahlstrand, 2.1.1.

4.2.0. The situation with the substantives is completely different.

4.2.1. The S/C substantive theme has two kinds of alternations: the zero vowel and the consonantal alternations.

4.2.1.1. In the ACA orthographic representation the word forms with the zero vowel realized must be processed as consisting of a special theme allomorph plus the ending, e g N sg momak-Ø, G pl momak-a, or G pl marak-a (marka).

4.2.1.2. Concerning the consonantal alternations, the S/C is both geographically and evolutionarily at a crossroads. The neighbouring Slovenian has the nominal theme completely without alternations and S/C's own Kajkavian dialects have a strong tendency towards this type of theme: this tendency weakens geographically towards the south and east.

In standard S/C, the alternating consonants are all the three velars, c and l, e g junak-Ø/junač-e/junac-i ..., zec-Ø/zeč-e/zeč-evi, sto/stol-a, nosilac/nosioc-a.

In the ACA orthographical representation even the conventional differences in writing the glide j must be processed as theme alternations (their phonemic character notwithstanding, viz Verdi-Ø, radi-o vs Verdij-a, radij-a.

4.2.1.3. Maybe as a consequence of the dialectal differences or of a general development tendency, both many old words and the innovations and borrowings respectively avoid the alternations or only use them as a possible variant in the S/C standard. Thus in N pl

masc there are, beside the normal alternation junak/junaci, codified forms like dečko-dečki, cucak-cucki, Bask-Baski, erg-ergi. Often forms on -i are avoided and replaced by the long plural, e g štrajk - štrajkovi, šah - šahovi (but padišah-padišasi), žig-žigovi. Very often the alternation velar/alveolar in vocative is avoided and in such cases the vocative ending -u instead of -e must be used, viz hercegu, riziku, cucku, Basku. There are plenty of of vacillations. No S/C dictionary or handbook codifies these instances systematically.

The alternations c/č and l/o have the same evolutionary weak position. Beside zec/zeče/zečevi there is plac/placu/placevi, fric/fricu/fricevi and kec/kečevi or kecovi, zečevi or zecovi. The l/o alternation in the theme "auslaut" is now only potential (sto, vo beside stol, vol), the classical grammar example palac is semantically differentiated into two lexemes palac/palca 'finger' and palac/paoca, paocu... 'spoke'.

Consequently the productive nomina agentis in /-l#c/ can even generalize one of the two theme allomorphs: thus words like nosioc, nosioca... have become very common.

4.2.2. Consequently, such a coherence (and, thus, mutual predictability) of the endings within a subparadigm, which we demonstrated with S/C verbs and adjectivals (4.1.1. above), does not exist in S/C substantives: it is most clear in the singular of the second (masculine) declension. Here, only the G and D-L sg endings are fully predictable and even in the accusative the N-A or G-A sg is used on defined premises (inanimacy vs animacy). The endings of vocative and instrumental are, however, only partly predictable. In V, the choice of the ending -e or -u is dependent on the fact whether there is a consonant alternation in the theme or not, cf oče-financu, momče-mačku, šahu: besides a vocative identical to the nominative is possible, this tendency increasing towards the northwest. The choice of the instrumental ending -om or -em is primarily dependent on theme auslaut (hard consonant and vowel, unless the glide is inserted, vs soft and c), secondly on the theme vocalism in the case of soft and c auslaut (kolačem, stricem vs ježom, kepecom, padežom or padežem): there are many vacillations and a secure codification is lacking. The same uncertainty concerning V and I exists even in plenty of words in -ar, -ir, -er (pekare/pekaru).

4.3. This often increases the number of theme allomorphs and would multiply the number of coherent ending series, such as were shown in 4.1.

For those reasons, another method was necessary when building up the STLX-1 (noun). Each ending has its own symbol already in the STLX; and in the ENLX those symbols are translated into the output language. Such symbols are e g "nso" = the ending -o in N sg, "nse" = the ending -e in N sg, "ns" = the ending -Ø in the N sg, "npi" = the ending -i in the N pl etc.

An example of an STLX-1 can be the lemma redak, see 3.2. above. This method allows us to register all vacillations as well, e g:

zec 1 ma zec ns gsa dsu asa vsu* iso npi gpa dpi ape xo*

zeč 1 ma zec vse xe

*)The forms with the theme levelling zecu and zecovi, zecova, zecovima, zecove.

The problems of vacillations in the STLX-1 reflect very well the levelling development processes within the S/C nouns.

5.0. In order to build up the STLXs quickly and without error a copying program KOPIA (see Ingemar Dahlstrand, 3.1.) was developed. This program can in principle utilize any lexeme from the STLXs as a copying model for a new lexeme (e g a new mrdati by the existing čitati or pričati or šetati ...), but in order to ensure the right models (and a little even for the linguistic reflex of systematicity) we have built up a copying model dictionary SMLX, containing model words for all productive as well as some unproductive flexion classes. This task was, again, quite easy for verbs and adjectivals, where the ending coherence within the subparadigms is sufficiently strong, but become almost a l'art pour l'art, when we tried to find all models for the masculines of the second declension.

In the following, two kinds of exhaustive descriptions of the masculines of the second declension are presented: an algorithm in key form and a tree diagram based on it. The terminal string of the tree diagram consists of 127 existing and 42 potential but not actually found model words for orthographic description of this substantive class, i e for the corresponding part of SMLX.

There is a good deal of uncertainty behind this description due to the lack of codification of S/C. In no dictionary or handbook is the existence of the long plural systematically stated. The same is true about the possibility of "normal" short plural (does a tenki exist , or only tenkovi, žigovi ...?), about the vocative

and about the instrumental. Our solution is based on the language competence of several educated native speakers, both Croats and Serbs.

ALGORITHM FOR THE MASCULINES OF THE 2nd DECLENSION

Abbreviations:

A	alternating consonant
IA	non-alternating consonant
ise	instr sg -em
iso	instr sg -om
ma	masculine animate
mi	masculine inanimate
ns	nom sg with zero ending
nso	nom sg with ending -o
vse	voc sg -e
vsu	voc sg -u
x	"long" plural
xo	in "long" plural the allomorph -ov-
xe	in "long" plural the allomorph -ev-

The numbers preceded by T will indicate the position of a model word inthe terminal string of the tree diagram (e g: Yes (=mi): žiri T1).

1. Without #?			
	Yes	goto	2
	No	goto	83
2. In N.sg. ns (= zero ending)?			
	Yes	goto	3
	No (= words like dečko, Mile)	goto	69
3. Theme without alternation?			
	Yes	goto	4
	No	goto	43
4. Theme finished in a vowel?			
	Yes	goto	5
	No	goto	10
5. Is this vowel -i?			
	Yes	goto	6
	No	goto	7

6. Mi?		
	Yes (= mi)	žiri T1
	No (= ma)	Verdi T2
7. Vocative in -u (vsu)?		
	Yes	goto 8
	No (=vs)	goto 9
8. Mi?		
	Yes (= mi)	bife T3
	No (= ma)	envoaje T4
9. (= vs) Mi?		
	Yes (= mi)	ragu T5
	No (= ma)	guru T6
10. Is the theme finished in another consonant than -c?		
	Yes	goto 11
	No (-c)	goto 28
11. Is the theme "hard"?		
	Yes	goto 12
	No (= finished in lj, nj, ć, đ, č, dž, š, ž, št, žd)	goto 21
12. Without grammatical (predesinential) suffix -in in sg.?		
	Yes (= without -in)	goto 13
	No (= with -in)	goto 20
13. The instrumental in -om and the vocative in -e (iso, vse)?		
	Yes	goto 14
	No (=many words in -ar, -er)	goto 17
14. Mi?		
	Yes (= mi)	goto 15
	No (= ma)	goto 16
15. (= mi) Without long plural?		
	Yes	izvor T7
	No	brod T8
16. (= ma) Without long plural?		
	Yes	direktor T9
	No	sin T10
17. Many words in -ar, -er) Mi?	Yes (= mi)	goto 18
	No (= ma)	goto 19

18. (= mi) Without long plural?		
	Yes (= vsu)	biser T11
	No	?
19. (= ma) Without long plural?		
	Yes	pekar T13
	No	?
20. (= words with -in) Without alternation in the theme?		
	Yes	srbin T15
	No	turčin T16
21. (= soft theme) Is the instrumental ending -em (ise)?		
	Yes	goto 22
	No	goto 25
22. (= ise) Mi?		
	Yes (= mi)	goto 23
	No (= ma)	goto 24
23. (= ise, mi) Without long plural?		
		kolač T17
	No	mač T18
24. (= ise, ma) Without long plural?		
	Yes	kovač T19
	No	kralj T20
25. (= iso) Mi?		
	Yes (= mi)	goto 26
	No (= ma)	goto 27
26. (= iso, mi) Without long plural?		
	Yes	padež T21
	No	kelj T22
27. (= iso, ma) Without long plural?		
	Yes	?
	No	?
28. (theme in -c) Without alternation -c/č?		
	Yes	goto 29
	No (= alternation -c/-č)	goto 36
29. (= theme in -c) Is		

the instrumental ending -em (ise)?

Yes (= ise)		goto 30
No (= iso)		goto 33

30. (= -c, ise) Mi?

Yes (= mi)		goto 31
No (= ma)		goto 32

31. (= mi) Without long plural?

Yes	?
No	plac T26

32. (= ma) Without long plural?

Yes	?
No	fric T28

33. (= -c, iso) Mi?

Yes (= mi)		goto 34
No (= ma)		goto 35

34. (= mi) Without long plural?

Yes	perec T29
No	?

35. (= ma) Without long plural?

Yes	čtec T31
No	?

36. (= theme in -c/-č) Is the instrumental ending -em (ise)?

Yes	goto 37
No	goto 40

37. (= -c/-č, ise) Mi?

Yes (= mi)		goto 38
No (= ma)		goto 40

38. (= mi) Without long plural?

Yes	ognjic T33
No	?

39. (= ma) Without long plural?

Yes	konjic T35
No	stric T36

40. (= -c/-č, iso) Mi?

Yes (= mi)		goto 41
No (= ma)		goto 42

41. (= mi) Without long

plural?

Yes	měsec T37
No	kec T38

42. (= ma) Without long plural?

Yes	kepec T39
No	zec T40

43. (= theme with alternation) Is the alternating consonant a velar?

Yes	goto 44
No	goto 66

44. Which velar is alternating?

If -k	goto 45
If -g	goto 52
If -h	goto 59

45. Is the alternation even in vocative?

Yes (= -k/č/c, vse)	goto 46
No (=k/c, vsu)	goto 49

46. (=-k/č/c, vse) Mi?

Yes (= mi)	goto 47
No (= ma)	goto 48

47. (= mi) Without long plural?

Yes	brzak T41
No	znak T42

48. (= ma) Without long plural?

Yes	junak T43
No	rak T44

49. (=-k/c, vsu) Mi?

Yes (= mi)	goto 50
No (= ma)	goto 51

50. (= mi) Without long plural?

Yes	rizik T45
No	štrejk T46

51. (= ma) Without long plural?

Yes	?
No	?

52. Is the alternation

even in vocative?

Yes	(=-g/ž/z, vse)	goto 53
No	(=-g/z, vsu)	goto 56

53. (=-g/ž/z, vse) Mi?

Yes	(= mi)	goto 54
No	(= ma)	goto 55

54. (= mi) Without long plural?

Yes	prtljag T49
No	brěg T50

55. (= ma) Without long plural?

Yes	herceg T51
No	vrag T52

56. (=g/z, vsu) Mi?

Yes	(= mi)	goto 57
No	(= ma)	goto 58

57. (= mi) Without long plural?

Yes	uteg T53
No	žig T54

58. (= ma) Without long plural?

Yes	mag T55
No	?

59. Is the alternation even in vocative?

Yes	(=-h/š/s, vse)	goto 60
No	(=-h/s, vsu)	goto 63

60. (=-h/š/s, vse) Mi?

Yes	(= mi)	goto 61
No	(= ma)	goto 62

61. (= mi) Without long plural?

Yes	trbuh T57
No	dah T58

62. (= ma) Without long plural?

Yes	vlah T59
No	duh T60

63. (=-h/s, vsu) Mi?

Yes	(= mi)	goto 64
No	(= ma)	goto 65

64. (= mi) Without long plural?

	Yes	tepih T61
	No	fah T62
65. (= ma) Without long plural?		
	Yes	padišah T63
	No	šah T64
66. (= alternating -l) Mi?		
	Yes (= mi)	goto 67
	No (= ma)	goto 68
67. (= mi) Without long plural?		
	Yes	razděo T65
	No	sto T66
68. (= ma) Without long plural?		
	Yes	anđeo T67
	No	vo T68
69. (= with a real ending) Is the real ending in N sg -o (nso)?		
	Yes	goto 70
	No	goto 79
70. (= nso) Is the theme without consonant alternation?		
	Yes	goto 71
	No (= many words in -ko, -go, -ho)	goto 75
71. Is the theme before -o (nso) finished in -i?		
	Yes (=-io)	goto 72
	No	goto 73
72. (=-io) Mi?		
	Yes (=mi)	radio T69
	No (=ma)	impresario T70
73. (= nso other than 72) Mi?		
	Yes (= mi)	goto 74
	No (= ma)	děčko T73
74. (= mi) Is the word masculine in plural?		
	Yes (= npi)	auto T71
	No (= npa)	dinamo T72
75. (= themes with alter-		

nation) Which consonant is alternating?

	If -k/c	goto 76
	If -g/z	goto 77
	If -h/z	goto 78

76. (= nso, -k/c) Mi?

	Yes (= mi)	riziko T74
	No (= ma)	rašćupanko T75

77. (= nso, -g/z) Mi?

	Yes (= mi)	embargo T76
	No (= ma)	flamingo T77

78. (= nso, -h/s) Mi?

	Yes (= mi)	?
	No (= ma)	Vlaho T79

79. (= nse) Without grammatical suffix -et- in casus obliqui?

	Yes	goto 80
	No (= gsa -eta)	Mile T80

80. (= nse) Is the instrumental ending -em?

	Yes (= ise)	goto 81
	No (= iso)	goto 82

81. (= nse, ise) Mi?

	Yes (= mi)	?
	No (= ma)	Blagoje T81

82. (= nse, iso) Mi?

	Yes (= mi) even ise	finale T82
	No (= ma)	Pavle T83

Beginning #

83. Is the zero vowel # followed by another consonant than -#c or -#k?

	Yes	goto 84
	No (= -ac or -ak)	goto 101

84. Is the # potential, i e non obligatory?

	Yes	goto 85
	No (= obligatory)	goto 88

85. (= not obligatory) Mi?

	Yes (= mi)	goto 86
	No (= ma)	goto 87

86. (= mi) Without long plural?
Yes — akcen(a)t T85
No — ?

87. (= ma) Without long plural?
Yes — docen(a)t T87
No — ?

88. (= obligatory #) Is the # followed by -l/o?
Yes (= -ao) — goto 89
No — goto 92

89. (= -#l/o) Mi?
Yes (= mi) — goto 90
No (= ma) — goto 91

90. (= mi) Without long plural?
Yes — nesmisao T89
No — kotao T90

91. (= ma) Without long plural?
Yes — dětao T91
No — pětao T92

92. Is the # followed by a hard consonant?
Yes (= hard theme) — goto 93
No (= soft theme) — goto 98

93. (= #, hard theme) Without voicing assimilation before #?
Yes (= no assimilation) — goto 94
No (= assimilation) — goto 97

94. (= #, hard theme) Mi?
Yes (= mi) — goto 95
No (= ma) — goto 96

95. (= mi) Without long plural?
Yes — srbizam T93
No — ritam T94

96. (= ma) Without long plural?
Yes — ministar T95
No — ?

97. (= voicing assimilation before #) Without

long plural?

Yes — hrbat T97
No — ?

98. (= #,soft theme) Mi?

Yes (= mi) — goto 99
No (= ma) — goto 100

99. (= mi) Without long plural?

Yes — ražanj T99
No — stupanj T100

100. (= ma) Without long plural?

Yes — ?
No — ?

101. Is the # followed by -c?

Yes (=-ac) — goto 102
No (=-ak) — goto 137

Beginning -#c

102. (=-ac) A non-alternating consonant (IA) before #c?

Yes — goto 103
No (= an alternating consonant before #c) — goto 110

103. (= IA#c) Mi?

Yes (= mi) — goto 104
No (= ma) — goto 105

104. (= mi) Without long plural?

Yes — lonac T103
No — šanac T104

105. (= ma) Without long plural?

Yes — němac T105
No — ?

106. Is the alternating consonant before # a non-obstruent (= -lac)?

Yes (=-lac) — goto 107
No (= -obstr.+ac) — goto 110

107. (=-lac) Mi?

Yes (= mi) — goto 108
No (= ma) — goto 109

108. (= mi) Without long plural?

	Yes	?
	No	goto 108x
108x. (= mi, long pl.) Is -lac preceded by an -o (= -olac)?		
	Yes	kolac T108
	No	palac T109
109. (= ma) Without long plural?		
	Yes	nosilac T110
	No	?
110. (= -A#c) Is the consonant alternation an assimilation (devoicing)?		
	Yes	goto 111
	No (= dental+ac)	goto 124
111. Which voiced obstruent precedes the -#c?		
	If b	goto 112
	If g	goto 115
	If ž	goto 118
	If đ	goto 121
112. (= -bac) Mi?		
	Yes (= mi)	goto 113
	No (= ma)	goto 114
113. (= mi) Without long plural?		
	Yes	zubac T112
	No	?
114. (= ma) Without long plural?		
	Yes	vrabac T114
	No	?
115. (= -gac) Mi?		
	Yes (= mi)	goto 116
	No (= ma)	goto 117
116. (= mi) Without long plural?		
	Yes	?
	No	?
117. (= ma) Without long plural?		
	Yes	bogac T118
	No	?
118. (= -žac) Mi?		
	Yes (= mi)	goto 119

	No (= ma)	goto 120
119. (= mi) Without long plural?		
	Yes	stožac T120
	No	?
120. (= ma) Without long plural?		
	Yes	lažac T122
	No	?
121. (= -đac) Mi?		
	Yes (= mi)	goto 122
122. (= mi) Without long plural?		
	Yes	gololeđac T124
	No	?
123. (= ma) Without long plural?		
	Yes	?
	No	?
124. Which dental consonant precedes the -#c?		
	If d	goto 125
	If t	goto 128
	If z	goto 131
	If s	goto 134
125. (= -dac) Mi?		
	Yes (= mi)	goto 126
	No (= ma)	goto 127
126. (= mi) Without long plural?		
	Yes	žaludac T128
	No	?
127. (= ma) Without long plural?		
	Yes	sudac T130
	No	?
128. (= -tac) Mi?		
	Yes (= mi)	goto 129
	No (= ma)	goto 130
129. (= mi) Without long plural?		
	Yes	prutac T132
	No	?
130. (= ma) Without long plural?		
	Yes	čatac T134
	No	otac T135

131. (= -zac) Mi?

Yes (= mi)		goto 132
No (= ma)		goto 133

132. (= mi) Without long plural?

Yes	obrazac T136
No	?

133. (= ma) Without long plural?

Yes	svizac T138
No	?

134. (= -sac) Without long plural?

Yes (= mi)		goto 135
No (= ma)		goto 136

135. (= mi) Without long plural?

Yes	kvasac T140
No	?

136. (= ma) Without long plural?

Yes	pisac T142
No	?

Beginning #k

137. (= -ak) Is -#k preceded by a non-alternating consonant (IA)?

Yes (= IA#k)		goto 138
No		goto 141

138. (= IA#k) Does the -k alternate with -g?

Yes (= -ak/g)		goto 139
No		goto 140

139. (= -zak/zg-) Mi?

Yes (= mi)		mozak T144
No (= ma)		drozak T145

140. (= IA#k) Mi*?

*Between words in -#k the long plural has not been found in accessible sources, but it is not excluded morphophonemically: its empty positions are not included in the algorithme, but they can be put in easily between the Mi-nodes and the terminal string.

Yes (= mi)	lěvak T146

	No (= ma)	momak T147
141. (= A#k) Is the consonant alternation before #k only a voicing assimilation (devoicing)?		
	Yes	goto 142
	No	goto 145
142. Which voiced consonant is alternating?		
	If b	goto 143
	If ž	goto 144
0143. (= -bak) Mi?		
	Yes (= mi)	šibak T148
	No (= ma)	golubak T149
144. (= -žak) Mi?		
	Yes (= mi)	?
	No (= ma)	žižak T151
145. Is the alternating consonant an affricate?		
	Yes (= -đak, -džak, -cak, -ćak, -čak)	goto 146
	No (= dent. +#k)	goto 154
146. (= affric. +#k) Is the affricate voiced?		
	Yes	goto 147
	No	goto 150
147. (= voiced affric. +#k) Which voiced affricate is alternating?		
	If đ	goto 148
	If dž	goto 149
148. (= -đak) Mi?		
	Yes (= mi)	omeđak T152
	No (= ma)	mlađak T153
149. (= -džak) Mi?		
	Yes (= mi)	kovrdžak T154
	No (= ma)	?
150. (= voiceless affric. +#k) Which voiceless affricate is alternating?		
	If c	goto 151
	If ć	goto 152
	If č	goto 153
151. (= -cak) Mi?		
	Yes (= mi)	?
	No (= ma)	cucak T157

152. (= -ćak) Mi?

Yes (= mi)	mućak T158
No (= ma)	dětićak T159

153. (= -čak) Mi?

Yes (= mi)	čičak T160
No (= ma)	mačak T161

154. (= dent.+#k) Is the alternating consonant a dental stop?

Yes (= dent.stop +#k)	goto 155
No (= dent.fricat.+#k)	goto 158

155. (= dental stop+#k) Which stop is alternating?

If d	goto 156
If t	goto 157

156. (= -dak) Mi*?

Yes (= mi)	redak T162
No (= ma)	medvědak T163

157. (= -tak) Mi*?

Yes (= mi)	zadatak T164
No (= ma)	patak T165

*Single very freqent words in -dak and -tak have vocative vse attached to a theme in -č and without the underlying dental, viz tetak teče. Normally however all lexemes sub 156 and 157 have the vocative vsu ending attached to the allomorph in -tk-.

158. (= dental fricative+#k) Which fricative is alternating?

If z	goto 159
If s	goto 160

159. (= -zak) Mi?

Yes (= mi)	svezak T166
No (= ma)	mrzak T167

160. (= -sak) Mi?

Yes (= mi)	vosak T168
No (= ma)	qusak T169

MASCULINES OF THE SECOND DECLENSION

A tree diagram to the algorithm on p. 132

Design M. Stankovski