

DISTINCTIVE FEATURES IN THE STRUCTURE OF ENGLISH AND SERBO-CROATIAN NOUN COMPOUNDS*

JELISAVETA MILOJEVIČ

University of Belgrade

1. Introduction

The aim of the research was to explore further the structure of English and Serbo-Croatian noun compounds, focusing on their inter-relations, in order to establish a set of criterial distinctive features on the basis of which the compounds are inter-related in terms of minimal pairs and to provide an algorithm for the identification of minimal pairs by computer.

The comprehensive literature on the subject was studied and critically assessed. A set of criterial distinctive features in the structure of English and Serb-Croatian noun compounds was hypothesized and the hypothesis tested by means of minimal pairs identified by computer. Conclusions concerning the internal structure of noun compounds and their systemic inter-relations were drawn. The algorithm for the identification of minimal pairs proved to be successful and had a further consequence of the possibility of treating ambiguities and synonymy in compounds in a computer precise way. A new model of English noun compounds was hypothesized and proved to be correct.

The compounds particularly chosen for the subject of the analysis were English and Serbo-Croatian hypotactic binomial noun compounds with the exponents of one verbal and one nominal element. A compound has been defined as a lexeme containing two or more potential stems and that has not subsequently been subjected to a derivational process. A binomial compound has been defined as a two-part compound. A noun compound is the one that can be assigned to the 'noun' form class. A binomial noun compound with an exponent of a verbal element is the one which is potentially related (in a generative i.e. synchronic way) to a verb

* I gratefully acknowledge the financial assistance of the British Scholarship Trust for Yugoslavs. I wish to thank Professor John McH. Sinclair for his comments and encouragement and Ljubomir Jankovič for his help with computer programming. The author alone is responsible for the views presented.

which has its exponent in the surface structure of a compound (e.g. *handwriting* in English and *blatobran* in Serbo-Croat). A binomial noun compound with an exponent of a nominal element is the one which is potentially related (in generative i.e. synchronic terms) to a noun which has its exponent in the surface structure of a compound (e.g. *washer-woman* in English and *ribolovac* in Serbo-Croat). The parts of a compound are not considered to have word status although they can be potentially related (in a generative and synchronic sense) to words. A hypotactic binomial compound is one in which the parts are in a hypotactic relation to each other (cf. hypotactic *water-carrier* vs. paratactic compound *carrier-pigeon*). The hypotactic relations between the parts of a binomial compound refer to hypotactic deep logical relations between an argument (or arguments) with respect to the verb (e.g. the temporal and agentive relation of the arguments with respect to the verb in *night-watch*).

After a comprehensive study of theoretical works, reference works, grammar books and dictionaries, a corpus of 2527 compounds (1526 English and 1001 Serbo-Croatian) was selected to be subjected to the analysis. The compounds were coded, so that they could be processed by computer, each code constituting a record and all records constituting a source data file. The computer program was written to perform several tasks, one of them being to identify and count minimal pairs. Algorithms for data processing were developed and were instrumental in proving the hypothesis. A comprehensive description of the way in which the coding was done and of the algorithm for the identification of ambiguities in compounds is presented in this paper. The model which was hypothesized was proven to be correct.

The explanation of the coding is presented in Section 2. and the description of the algorithm in Section 4. The computer program was written in BASIC and the computer used for processing was an IBM PC. Distinctive features in the structure of English and Serbo-Croatian noun compounds are presented in Section 3. Lists of minimal pairs in both languages are given in Section 5. Discussion and concluding remarks follow in Section 6.

2. Explanation of the code

The record consists of 16 fields with the following syntax:

F1_F2_F3...F16

Each field has the syntax:

x1x2x3...xn

where x is a single character sign and n an integer number. The syntax of each field is presented in the text which follows.

Field F1 has the following syntax:

a1a2a3...a18

where a1 to a18 are characters of a binomial noun compound with a verbal element.

Field F2 has the following syntax:

b1b2

where b1 to b2 denote the source (dictionary, TV, newspapers, spoken language, etc.).

Field F3 has the following syntax:

c1c2c3...c10

where c1 to c10 are characters of the verb with which the verbal element is associated.

Field F4 has the following syntax:

d1

where d1 is a character representing part of the compound in which the verbal element occurs. The following notation is applied:

- d1=1 – if the verbal element occurs in the first part of the compound;
- d2=2 – if the verbal element occurs in the second part of the compound.

Field F5 has the following syntax:

e1e2e3...e9

where e1 to e9 are characters of the verb with which the verbal element is associated.

Field F6 has the following syntax:

f1

where f1 is a character representing part of the compound in which the noun element occurs. The following notation is applied:

- f1=1 – if the noun element occurs in the first part of the compound;
- f1=2 – if the noun element occurs in the second part of the compound.

Field F7 has the following syntax:

g1g2g3...g5

where g1 to g5 are characters of a linking element in the first part of the compound. The following notation is applied:

- g1 to g5 = a,b,c,d,e...z – if there is a linking element in the first part of the compound;
- g1=0 – if there is no linking element in the first part of the compound.

Field F8 has the following syntax:

h1

The following notation is applied:

- $h1=0$ – if there is no linking element in the first part of the compound;
- $h1=1$ – if the linking element follows the verbal element;
- $h1=2$ – if the linking element does not follow the verbal element.

Field F9 has the following syntax:

$i1i2i3...i5$

where $i1$ to $i5$ are characters of a bound form in the second part of the compound.

The following notation is applied:

- $i1$ to $i5 = a,b,c,d,e...z$ – if there is a bound form in the second part of the compound;
- $i1 = 0$ – if there is no bound form in the second part of the compound.

Field F10 has the following syntax:

$j1$

The following notation is applied:

- $j1=0$ – if there are no bound forms in the second part of the compound;
- $j1=1$ – if a bound form follows the verbal element;
- $j1=2$ – if a bound form does not follow the verbal element.

Field F11 has the following syntax:

$k1$

where $k1$ is a character denoting agentivity or nonagentivity of the verb associated with the verbal element in the compound. The following notation is applied:

- $k1=1$ – if the verb with which the verbal element is associated is agentive;
- $k1=2$ – if the verb with which the verbal element is associated is nonagentive.

Field F12 has the following syntax:

$l1$

where $l1$ is a character denoting the number of overt relations. The following notation is applied:

- $l1=1$ – if there is one overt relation;
- $l2=2$ – if there are two overt relations.

Field F13 has the following syntax:

$m1m2m3...m10$

where $m1$ to $m10$ are the characters denoting agentivity ($m1$); marker ($m2$); overt

relation ($m3$ to $m5$); overt relation obligatory/nonobligatory ($m6$); (overt relation) ($m7$ to $m9$); (overt relation obligatory/nonobligatory) ($m10$). The following notation is applied;

- $m1=1$ – if the verb with which the verbal element is associated is agentive;
- $m1=2$ – if the verb with which the verbal element is associated is nonagentive;
- $m2=0$ – if there is no marker;
- $m2=1$ – ‘purpose’ marker;
- $m2=2$ – ‘habitual’ marker;
- $m2=3$ – ‘habitual and professional’ marker;
- $m2=4$ – ‘species’ marker;
- $m3$ to $m5=R01$ – locative;
- $m3$ to $m5=R02$ – instrumental;
- $m3$ to $m5=R03$ – possessor;
- $m3$ to $m5=R04$ – patient;
- $m3$ to $m5=R05$ – factitive;
- $m3$ to $m5=R06$ – agentive;
- $m3$ to $m5=R07$ – objective;
- $m3$ to $m5=R08$ – source;
- $m3$ to $m5=R09$ – temporal;
- $m3$ to $m5=R10$ – goal;
- $m3$ to $m5=R11$ – commitative;
- $m3$ to $m5=R12$ – possessed;
- $m6=1$ – nonobligatory;
- $m6=2$ – obligatory;
- $m7$ to $m9$ – explanation as in $m3$ to $m5$;
- $m10$ – explanation as in $m6$.

Field F14 has the following syntax:

$n1n2n3...n10$

where $n1$ to $n10$ are characters denoting overt relation(s) and the verb with which the verbal element is associated in the order of representation in the form of the compound; marker. Notation applied is the same as in F13 except for characters V and M put in front of the characters denoting agentivity/nonagentivity and the type of the marker respectively.

Field F15 has the following syntax:

$o1$

where $o1$ is a character denoting whether the argument structure of the compound is or is not satisfied within the compound. The following notation is applied:

o1=0 – if the argument structure is not satisfied within the compound;
o1=1 – if the argument structure is satisfied within the compound.

Field F16 has the following syntax:

p1p2p3...p6

where p1 to p3 and p4 to p6 are characters denoting covert relations. The following notation is applied:

p1 to p3 = P00	– if there are no covert relations;
p1 to p3 (and or p4 to p6) = P01	– locative;
p1 to p3 (and or p4 to p6) = P02	– instrumental;
p1 to p3 (and or p4 to p6) = P03	– possessor;
p1 to p3 (and or p4 to p6) = P04	– patient;
p1 to p3 (and or p4 to p6) = P05	– factitive;
p1 to p3 (and or p4 to p6) = P06	– agentive;
p1 to p3 (and or p4 to p6) = P07	– objective;
p1 to p3 (and or p4 to p6) = P08	– source;
p1 to p3 (and or p4 to p6) = P09	– temporal;
p1 to p3 (and or p4 to p6) = P010	– goal;
p1 to p3 (and or p4 to p6) = P011	– commitative;
p1 to p3 (and or p4 to p6) = P012	– possessed.

3. Distinctive features

The system of distinctive features that has been hypothesized and tested has not been taken from any previous research by other authors.

For a general survey of the distinctive features of compounds the reader is referred to the previous section. The features which characterize both overt and covert relations as well as the relationship between the relations and relational features and further between the relations and lexical features of the verb with which the participating verbal element in the compound is associated are summarized in Table 1.

The relations are to be understood as deep logical relations between an argument (or arguments) and the verb. They are defined by the total of the relational features marked by (+) (see Table 1.). Illustrations are given in Tables 2 and 3.

Table I. Relations, lexical and relational features

LFV RELATIONS		RELATIONAL FEATURES											
A	NA	CLR	CLD	CAU	EFF	SRC	GOA	ACT	PAS	AFS	AFD	LOC	TIM
+	–	AGENTIVE	+	–	+	–	+	–	+	–	+	–	–
+	–	INSTRUMENTAL	–	+	+	–	+	–	–	+	–	–	–
–	+		–	+	+	–	–	–	–	+	–	–	–
			+	–	+	–	+	–	–	+	+	–	–
+	–	OBJECTIVE	–	+	–	–	–	+	–	+	–	+	–
+	–	FACTITIVE	–	+	–	+	–	+	–	+	–	–	–
–	+	PATIENT	–	+	–	–	–	+	–	+	–	+	–
+	–	SOURCE	–	–	–	–	+	–	–	+	–	–	–
–	+												
+	–	GOAL	–	–	–	–	–	+	–	+	–	–	–
–	+	POSSESSOR	+	–	–	–	–	–	–	+	–	–	–
–	+	POSSESSED	–	+	–	–	–	–	–	+	–	–	–
+	–	COMMITATIVE	–	–	–	–	–	+	+	–	–	–	–
+	–	LOCATIVE	–	–	–	–	–	–	–	–	–	–	+
–	+												
+	–	TEMPORAL	–	–	–	–	–	–	–	–	–	–	+
–	+												

LFV = Lexical feature of the verb	GOA = Goal
A = Agentive	ACT = Active
NA = Nonagentive	PAS = Passive
CLR = Controller	AFS = Affects
CLD = Controlled	AFD = Affected
CAU = Cause	LOC = Location
EFF = Effect	TIM = Time
SRC = Source	

Table II. Relations, lexical and relational features illustrated by means of minimal pairs in English

RELATIONS		
Agent	+	wine-taster
Agent	–	wine-cooler
Agent	+	candle-lighter
Agent	–	candle-lighter
Instrument 1	+	walking-part
Instrument 1	–	walking-shoe
Instrument 1	+	mercy-killing
Instrument 1	–	time-killing
Instrument 2	+	ear-picker
Instrument 2	–	hop-picker
Instrument 2	+	sword-dance
Instrument 2	–	ice-dance

Instrument 3	+	doorstopper
Instrument 3	-	door-keeper
Instrument 3	+	grinding-wheel
Instrument 3	-	driving-wheel
Object	+	housekeeper
Object	-	house-builder
Object	+	call-girl
Object	-	call-boy
Factitive	+	shipbuilder
Factitive	-	ship-breaker
Factitive	+	road-maker
Factitive	-	road-scraper
Patient	+	ice fall
Patient	-	windfall
Patient	+	ice-drift
Patient	-	ice-dance
Source	+	reading-book
Source	-	reading-lamp
Source	+	prison-breaker
Source	-	housebreaker
Goal	+	housebreaker
Goal	-	prison-breaker
Goal	+	sunshine
Goal	-	sunworship
Possessor	+	ship-owner
Possessor	-	shipbuilder
Possessor	+	householder
Possessor	-	housekeeper
Possessed	+	householder
Possessed	-	house-cleaner
Possessed	+	house-owner
Possessed	-	housekeeper
Commitative	+	work-mate
Commitative	-	workman
Commitative	+	playmate
Commitative	-	plaything
Locative	+	playground
Locative	-	playtime
Locative	+	writing-paper
Locative	-	blotting-paper
Temporal	+	wash-day
Temporal	-	wash-house
Temporal	+	playtime
Temporal	-	kill-time

LEXICAL FEATURES OF THE VERB		
Agentive	+	folding-machine
Agentive	-	folding-door
Agentive	+	shoe-shine
Agentive	-	sunshine
RELATIONAL FEATURES		
Controllor	+	call-boy
Controllor	-	call-girl
Controllor	+	potato picker
Controllor	-	ear-picker
Controlled	+	dancing shoe
Controlled	-	dancing-girl
Controlled	+	walking-stick
Controlled	-	walking-lady
Cause	+	sneeze-wood
Cause	-	drift-wood
Cause	+	windfall
Cause	-	rainfall
Effect	+	house-builder
Effect	-	house-cleaner
Effect	+	cheese maker
Effect	-	cheese-slicer
Source	+	prison-breaker
Source	-	house-breaker
Source	+	sunshine
Source	-	sunset
Goal	+	prize-fight
Goal	-	sea-fight
Goal	+	shop-breaker
Goal	-	prison-breaker
Active	+	call-bird
Active	-	call-girl
Active	+	dancing bear
Active	-	dancing-shoe
Passive	+	catchfly
Passive	-	catchweed
Passive	+	sucking-bottle
Passive	-	sucking-pig
Affector	+	drawing-machine
Affector	-	drawing-bridge
Affector	+	sleeping-pill
Affector	-	sleeping-bag
Affected	+	ice-breaker
Affected	-	ice-skater

Affected	+	stage-driver
Affected	-	stage-player
Location	+	sea-fight
Location	-	bull-fight
Location	+	heartache
Location	-	heart-break
Time	+	daydream
Time	-	daybreak
Time	+	feeding-time
Time	-	feeding-bottle
MARKERS AS FEATURES		
Purpose	+	bottle-washer
Purpose	-	bottle-washer
Purpose	+	kill-time
Purpose	-	kill-joy
Habitual	+	dish-washer
Habitual	-	dish-washer
Habitual	+	bird-watcher
Habitual	-	bird trainer
Habitual & Professional	+	window-washer
Habitual & Professional	-	bottle-washer
Habitual & Professional	+	bird trainer
Habitual & Professional	-	bird-watcher
Species	+	grave-digger
Species	-	grave-digger
Species	+	fly-catcher
Species	-	fly-catcher

Table III. Relations, lexical and relational features illustrated by means of minimal pairs in Serbo-croat

RELATIONS		
Agent	+	kitolovac
Agent	-	kitolovac
Agent	+	ribolovac
Agent	-	kitolovac
Instrument 1	+	vetrobran
Instrument 1	-	domobran
Instrument 1	+	blatobran
Instrument 1	-	biljobran
Instrument 2	+	kitolovac
Instrument 2	-	ribolovac
Instrument 2	+	rukomet
Instrument 2	-	vatromet
Instrument 3	+	ledolomac
Instrument 3	-	brakolomac

Instrument 3	+	biljobran
Instrument 3	-	domobran
Object	+	domobran
Object	-	suncobran
Object	+	gasovod
Object	-	cevovod
Factitive	+	zakonopisac
Factitive	-	zakonodavac
Factitive	+	zlotvor
Factitive	-	zloslut
Patient	+	mrzismet
Patient	-	beismet
Patient	+	miroljubac
Patient	-	mirotvorac
Source	+	smrdibuba
Source	-	strizibuba
Source	+	beismet
Source	-	mrzismet
Goal	+	bogomoljac
Goal	-	bogobojac
Goal	+	mrzismet
Goal	-	beismet
Possessor	+	zemljodržac
Possessor	-	zemljodelac
Possessor	+	zemljoposednik
Possessor	-	zemljoradnik
Possessed	+	zemljodržac
Possessed	-	zemljodelac
Possessed	+	vlastodržac
Possessed	-	silodržac
Comitative	+	pevidrug
Comitative	-	0
Comitative	+	plačidrug
Comitative	-	0
Locative	+	čelovodja
Locative	-	četovodja
Locative	+	sudopera
Locative	-	sudopera
Temporal	+	spomen-dan
Temporal	-	spomen-ploča
Temporal	+	listopad
Temporal	-	vodopad
LEXICAL FEATURES OF THE VERB		
Agentive	+	zemljopis
Agentive	-	zemljotres

Agentive	+	bežisvet
Agentive	-	mrzismet
RELATIONAL FEATURES		
Controllor	+	drvocep
Controllor	-	drvocep
Controllor	+	crvotoč
Controllor	-	vinotoč
Controlled	+	knjigotoč
Controlled	-	crvotoč
Controlled	+	vlastodržac
Controlled	-	silodržac
Cause	+	blatobran
Cause	-	biljobran
Cause	+	jadomor
Cause	-	čedomor
Effect	+	dobrotvor
Effect	-	rukotvor
Effect	+	čudotvor
Effect	-	umotvor
Source	+	smrdibuba
Source	-	strizibuba
Source	+	bežisvet
Source	-	mrzismet
Goal	+	zemljovod
Goal	-	cevovod
Goal	+	bogoljubac
Goal	-	bogobojac
Active	+	bogopsovac
Active	-	bogomrzac
Active	+	drvogríz
Active	-	drvoljub
Passive	+	vlastodržac
Passive	-	vlastodavac
Passive	+	miroljubac
Passive	-	mirotvorac
Affector	+	jedibaba
Affector	-	smrdibaba
Affector	+	tužibaba
Affector	-	visibaba
Affected	+	zemljotres
Affected	-	zemljovod
Affected	+	konjoder
Affected	-	konjoljub
Location	+	bogomolja
Location	-	bogomolja

Location	+	čelovodja
Location	-	četovodja
Time	+	listopad
Time	-	vodopad
Time	+	kolovoz
Time	-	kolovoz
MARKERS AS FEATURES		
Purpose	+	mamipara
Purpose	-	mamipara
Purpose	+	ledolomac
Purpose	-	brakolomac
Habitual	+	deriklupa
Habitual	-	derikoža
Habitual	+	vodonoša
Habitual	-	glavonoša
Habitual & Professional	+	derikoža
Habitual & Professional	-	derikoža
Habitual & Professional	+	pismonoša
Habitual & Professional	-	bakciloniša
Species	+	kažiput
Species	-	kažiprst
Species	+	biljojed
Species	-	žabojed

4. Algorithm for identifying minimal pairs by computer

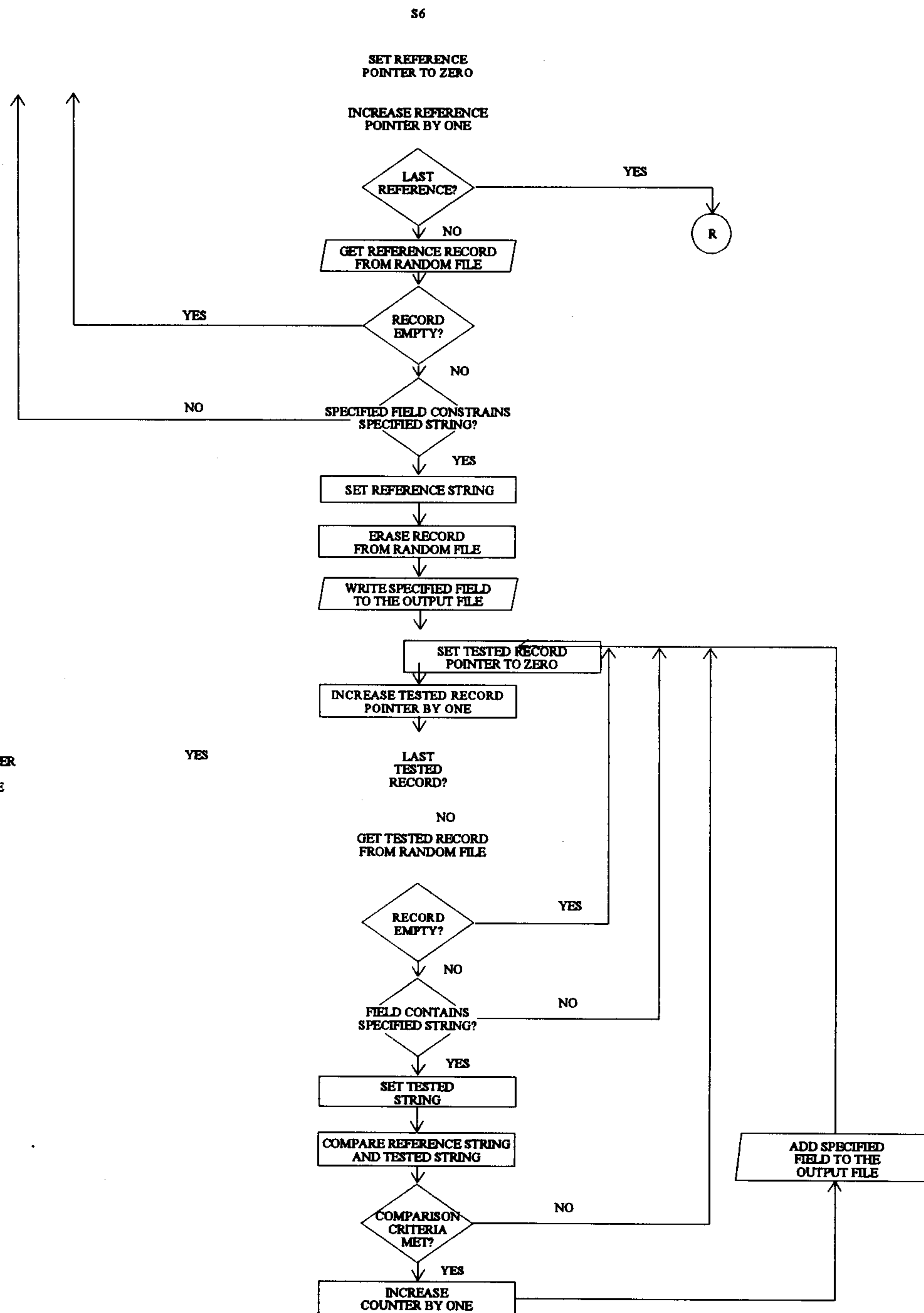
The program searches the source data file for records with codes specified in the input file (job card) and sets the first of those for reference. At the same time it deletes the record from the source data file. After that it starts searching the source data file from the beginning for the combination of codes (specified in the input data file) which matches that of the reference. Every match results in incrementing the counter. Having searched the whole source data file for records which match the reference, the program sets the new reference and repeats the whole procedure. The flow chart of the algorithm for identifying minimal pairs is presented in Fig. 1.

The description of the input data file (job card) is as follows:

Line No.	Parameter	Description
1:	fin	Input source data file name
2:	fout	Output data file name
3:	n	Maximum number of records
4:	t1	Task selection string
5:	t2	Task specification string

TASK 1 SPECIFICATION has the following syntax:

F1,F21,STRING1\$,F31,F32,F33,F34,F35,F41,F42



where

- 1 if the output is the F1 string;
- F21=4 if the specified characteristic is in F4;
- STRING1\$=1 if the specified characteristic is 1;
- F31=3 if the reference record and the matching one are to be the same with regards to F3;
- F32=7 if the reference record and the matching one are to be the same with regards to F7;
- F33=8 if the reference record and the matching one are to be the same with regards to F8;
- F34=9 if the reference record and the matching one are to be the same with regards to F9;
- F35=10 if the reference record and the matching one are to be the same with regards to f10;
- F41=14 if the reference record and the matching one are to be different with regards to F14;
- F42=16 if the reference record and the matching one are to be different with regards to F16.

TASK 2 SPECIFICATION has the following syntax:

F1,F21,STRING1\$,F31,F32,F33,F34,F35,F41,F42

where

- F1=1 if the output is the F1 string;
- F21=4 if the specified characteristic is in F4;
- STRING1\$=2 if the specified characteristic is 2;
- F31=3 if the reference record and the matching one are to be the same with regards to F3;
- F32=7 if the reference record and the matching one are to be the same with regards to F7;
- F33=8 if the reference record and the matching one are to be the same with regards to F8;
- F34=9 if the reference record and the matching one are to be the same with regards to F9;
- F35=10 if the reference record and the matching one are to be the same with regards to f10;
- F41=14 if the reference record and the matching one are to be different with regards to F14;
- F42=16 if the reference record and the matching one are to be different with regards to F16.

TASK 3 SPECIFICATION has the following syntax:

F1,F21,STRING1\$,F31,F32,F33,F34,F35,F41,F42

where

F1=1	if the output is the F1 string;
F21=4	if the specified characteristic is in F4;
STRING1\$=2	if the specified characteristic in F4 is 2;
F31=5	if the reference record and the matching one are to be the same with regards to F5;
F32=7	if the reference record and the matching one are to be the same with regards to F7;
F33=8	if the reference record and the matching one are to be the same with regards to F8;
F34=9	if the reference record and the matching one are to be the same with regards to F9;
F35=10	if the reference record and the matching one are to be the same with regards to f10;
F41=14	if the reference record and the matching one are to be different with regards to F14;
F42=16	if the reference record and the matching one are to be different with regards to F16.

TASK 4 SPECIFICATION has the following syntax:

F1,F21,STRING1\$,F31,F32,F33,F34,F35,F41,F42

where

F1=1	if the output is the F1 string;
F21=4	if the specified characteristic is in F4;
STRING1\$=1	if the specified characteristic is 1;
F31=5	if the reference record and the matching one are to be the same with regards to F5;
F32=7	if the reference record and the matching one are to be the same with regards to F7;
F33=8	if the reference record and the matching one are to be the same with regards to F8;
F34=9	if the reference record and the matching one are to be the same with regards to F9;
F35=10	if the reference record and the matching one are to be the same with regards to f10;
F41=14	if the reference record and the matching one are to be different with regards to F14;
F42=16	if the reference record and the matching one are to be different with regards to F16.

5. Lists of minimal pairs

There are four lists of minimal pairs (English examples) and four lists of minimal pairs (Serbo-Croatian examples) which are TASK 1, TASK 2, TASK 3 and

TASK 4 output. The lists are given in Tables 4 to 11. They are not comprehensive and only a small part of the output is presented owing to lack of space.

Minimal pairs are presented in groups followed by a number corresponding to the number of pairs. The first item in the group forms a pair with any other in the group.

Table IV. List of minimal pairs - Task 1 output (extracts)

answerphone COUNT= 0	catch-phrase catch-pit COUNT= 6	hunting-ant hunting-cog hunting-field hunting-ground hunting-party COUNT= 4	playing-card playing-field COUNT= 1
bake-house bake-stone COUNT= 1	check-point check-time COUNT= 1	hunting-field hunting-party hunting-spider COUNT= 2	reading glove reading-book reading-desk reading-room COUNT= 3
baking-powder COUNT= 0	chewing-gum COUNT= 0	kill-devil kill-time COUNT= 1	riding-boot riding-horse riding-school COUNT= 2
breakneck breakwater break-wind COUNT= 2	cooking-chesse cooking-oil COUNT= 1	kill-joy kill-time COUNT= 1	scrubbing-board scrubbing-brush COUNT= 1
breakwater break-jaw COUNT= 1	climbing plant climbing-boot COUNT= 1	looking-glass COUNT= 0	sucking-bottle sucking-pig COUNT= 1
call button call-bird callbox call-boy call-day call-girl COUNT= 5	cutpurse cutwater cut-worm COUNT= 2	magnifying glass COUNT= 0	suck-bottle suck-egg suck-fish suck-thumb COUNT= 3
call-bell call-bird call-box call-boy call-day call-girl COUNT= 5	dancing bear dancing shoe dancing-girl dancing-partner COUNT= 3	playboy playcentre playfellow playground playmate playroom plaything playtime play-day	swingboat swingtree swing-bridge swing-door COUNT= 3
catchfly catchpenny catchweed catchword catch-basin	grinding-wheel COUNT= 0	play-school COUNT= 9	
	grindstone COUNT= 0		

Table V. List of minimal pairs - Task 2 output (extracts)

ant killer	dirt-eater	cock-fight	water-finder
cane-killer	fire-eater	pillow-fight	witch-finder
joy-killer	fish-eater	prize-fight	COUNT= 5
king-killer	frog-eater	sea-fight	fire break
lady-killer	lotus-eater	shadow-fight	jailbreak
time-killer	man-eater	COUNT= 5	wind-break
COUNT= 5	toad-eater	bullfighter	COUNT= 2
arrow-maker	COUNT= 7	bush-fighter	hand-worker
coffee maker	blood-poisoning	prize-fighter	night-warker
holidaymaker	lead-poisoning	prize-fighter	COUNT= 1
image-maker	COUNT= 1	COUNT= 3	hat stand
law-maker	bone-breaker	butter-cutter	headstand
match-maker	circuit-breaker	diamond-cutter	shoulder stand
mischief-maker	diamond-breaker	gem-cutter	COUNT= 2
money-maker	housebreaker	glass-cutter	horse keeper
money-maker	ice-breaker	leaf-cutter	time-keeper
peacemaker	nut breaker	letter-cutter	COUNT= 1
rain-maker	prison-breaker	mable-cutter	house-cleaner
sandwich maker	safebreaker	plant-cutter	vacuum-cleaner
trouble-maker	shell-breaker	type-cutter	COUNT= 1
wax-maker	ship-breaker	wheel-cutter	ice fall
way-maker	shop-breaker	wood-cutter	windfall
COUNT= 14	stone-breaker	COUNT= 10	COUNT= 1
baby-walker	sword-breaker	egg-beater	insect-feeder
fire-walker	COUNT= 12	gold-beater	paper-feeder
nightwalker	bone-setting	COUNT= 1	COUNT= 1
rope-walker	sun-setting	eye-wash	shoeshine
shop-walker	COUNT= 1	gold-wash	sky shine
space walker	bone-shaker	rain-wash	starshine
street-walker	cocktail-shaker	COUNT= 2	sunshine
wire-walker	hand-shaker	fault-finder	COUNT= 3
COUNT= 7	saltshaker	gold-finder	
back-scratcher	COUNT= 3	key-finder	
COUNT= 0	bee-eater	position-finder	

Table VI. List of minimal pairs - Task 3 output (extracts)

breadwinner	coffee maker	flyswatter	rope-walker
bread maker	coffee whitener	fly killer	COUNT= 2
COUNT= 1	coffee-drinker	fly-catcher	sunburn
bribe-giver	coffee-grinder	fly-catcher	sunrise
bribe-taker	coffee-grower	fly-catcher	sunset
COUNT= 1	coffee-producer	fly-fisher	sunshine
brickmaker	COUNT= 5	COUNT= 5	sun-worship
brick pointer	climbing-boot	gas-fitter	COUNT= 4
brick-burner	coffee-growing	gas-heater	toothache
brick-setter	coffee-producing	gas-washer	toothpick
COUNT= 3	COUNT= 1	COUNT 2	toothwash
butter maker	bathing-machine	house-building	COUNT= 2
butter-cooler	brushing-machine	house-cleaning	water-breather
butter-cutter	composing-machine	house-owning	water-carrier
COUNT= 2	drawing-machine	house-warming	water-carrier
candle-lighter	drilling-machine	COUNT= 3	water-cooler
candle-lighter	folding-machine	ice-dance	water-drinker
COUNT= 1	mixing-machine	ice-drift	water-finder
cave-diver	printing-machine	ice-fall	water-softener
cave-dweller	rolling-machine	ice-quake	COUNT= 6
COUNT= 1	sewing-machine	COUNT= 3	wire-dancer
chair maker	stamping-machine	peacemaker	wire-drawer
chair mender	thrashing-machine	peace-breaker	wire-puller
chair-warmer	washing-machine	peace-keeper	wire-walker
COUNT= 2	weighing-machine	peace-lover	COUNT= 3
cheese cutter	wringing-machine	peace-seeker	word-building
cheese maker	COUNT= 14	COUNT= 4	word-painting
cheese taster	blazing-star	road-maker	word-splitting
cheese trier	falling star	road-mender	COUNT= 2
COUNT= 3	shooting-star	road-runner	
cock-crowing	COUNT= 2	road-scraper	
cock-fighting	riding-boot	COUNT= 3	
COUNT= 1	walking boot	rope-dancer	
	COUNT= 2	rope-maker	

Table VII. List of minimal pairs - Task 4 output (extracts)

blotting-paper	dining room	dwelling-place	surfing-board
drawing-paper	living room	eating-place	washing-board
printing-paper	resting-room	hiding-place	COUNT= 1
tracing paper	sitting-room	landing-place	surf-boat
wrapping-paper	smoking-room	sticking-place	swingboat
writing-paper	standing-room	watering-place	COUNT= 1
COUNT= 5	waiting-room	COUNT= 5	swimming-pool
call button	writing-room	looking-glass	watering-pool
push-button	COUNT= 8	magnifying glass	COUNT= 1
COUNT= 1	climbing-boot	reading-glass	throwing-stick
call-box	riding-boot	COUNT= 2	walking-stick
sneeze-box	walking boot	ironing-board	COUNT= 1
spit-box	COUNT= 2	running-board	tracing-paper
workbox	dancing shoe	scrubbing-board	writing-paper
COUNT= 3	jogging shoe	surfing-board	COUNT= 1
carving fork	running shoe	washing-board	COUNT= 4
toasting-fork	walking-shoe	COUNT= 4	walking-day
tuning-fork	COUNT= 3	scratch board	washing-day
COUNT= 2	diving board	skateboard	COUNT= 1
catchweed	draining board	springboard	wash-day
choke-weed	drawing-board	surf-board	work-day
driftweed	ironing-board	COUNT= 3	COUNT= 1
sneeze-weed	scrubbing-board	sneeze-wood	wash-woman
stinkweed	washing-board	stink-wood	work-woman
COUNT= 4	COUNT= 5	COUNT= 1	COUNT= 1
change-room	dwelling-house	spinning-wheel	watchman
rest-room	eating-house	steering-wheel	workman
smoke-room	gambling house	COUNT= 1	COUNT= 1
workroom	printing-house	stumbling-stone	
COUNT= 3	washing-house	wishing-stone	
changing-room	COUNT= 4	COUNT= 1	

Table VIII. List of minimal pairs - Task 1 output (extracts)

deriguša	kažiput	spomen-dan	vrtiguz
deriklupa	kaživetar	COUNT= 0	COUNT= 0
derikoža	COUNT= 3	strizibuba	vrtikapa
COUNT= 2		COUNT= 0	COUNT= 0
gulibrada	skočibuba	tarikamen	vrtoglavica
gulikoža	skočidevojka	COUNT= 0	COUNT= 0
gulikoža	COUNT= 1	tecikuča	
COUNT= 2		COUNT= 0	
dažiprst	smrdibuba	visibaba	
kažiput	smrdivrana	COUNT= 0	
	COUNT= 1		

Table IX. List of minimal pairs - Task 2 output (extracts)

badnjonosac	bakropisac	blatobran	bogomoljka
knjigonosac	basnopisac	domobran	bogomoljka
mačonosac	člankopisac	domobran	COUNT= 1
minonosac	firmopisac	kamenobran	bubolovac
vodonosac	ikonopisac	kišobran	buvolovac
COUNT= 4	stihopisac	ledobran	kitolovac
bakcilonoš	zakonopisac	snegobran	klasolovac
glasonoš	COUNT= 6	suncobran	minolovac
glavonoša	bakrorez	vatrobran	misolovac
kljuconoša	bakrorez	vetrobran	mravolovac
njigonoša	drvorez	COUNT= 10	pticolovac
kopljonoša	ikonorez	biljojed	tunolovac
maconoša	ledorez	mačkojed	COUNT= 8
mlekonoša	zlatorez	zemljojed	četovodja
pismonoša	COUNT= 5	zmijojed	hrtovodja
plodonoša	basnopisac	žabojed	kolovodja
postonoša	kamenopisac	COUNT= 4	medovodja
rogonoša	zakonopisac	bogoljub	slepcovodja
sporonoša	COUNT= 2	drvoljub	volovodja
toplonoša	basnotvorac	vodoljub	COUNT= 5
venconoša	čudotvorac	COUNT= 2	čudotvorac
COUNT= 14	dobrotvorac	bogoljubac	pesmotvorac
bakropis	jezikotvorac	vodoljubac	rukotvorac
ikonopis	mirotvorac	COUNT= 1	stihotvorac
putopis	rukotvorac	bogomolja	COUNT= 3
rukopis	zakonotvorac	bogomolja	dobrotvor
šakopis	COUNT= 6	bogomolja	rukotvor
zemljopis	biljobran	COUNT= 2	umotvor
COUNT= 5			COUNT= 2

drvomor	muvomor	rukopis	kamenobran
jadomor	COUNT= 5	šakopis	ledjobran
klicomor		zemljopis	mostobran
kozomor	ikonopis	COUNT= 4	rukobran
mišomor	putopis		COUNT= 3

Table X. List of minimal pairs - Task 3 output (extracts)

bakarorez	bogobojac	drvocep	kišobran
bakarorez	bogohulac	drvodelj	kišomer
bakropis	bogoljubac	drvoljub	COUNT= 1
bakrorez	bogomoljac	drvomer	krvolok
COUNT= 3	bogomrzac	drvorez	krvotok
	bogožalac	drvorez	COUNT= 1
bakropisac	COUNT= 4	drvosek	poslodavac
bakrorezac	bogohvalka	drvosek	posloprimac
COUNT= 1	bogomoljka	COUNT= 9	COUNT= 1
basnoljubac	bogomoljka	glavobolja	rukodelac
basnopsisac	bogomoljka	glavonoša	rukotvorac
basnotvorac	COUNT= 3	glavoseča	rukovodilac
COUNT= 2		glavoseča	COUNT= 2
biljober	brodobran	COUNT= 3	
biljobran	brodolom	kamenolomac	
biljojed	COUNT= 1	kamenopisac	
biljožder	drvocep	COUNT= 1	
COUNT= 3	drvocep		

Table XI. List of minimal pairs - Task 3 output (extracts)

kradikesa	mlatipara	pevidrug	gladibrk
vežikesa	COUNT= 2	plačidrug	COUNT= 0
COUNT= 1	mastibrk	COUNT= 1	gonivetar
	COUNT= 0	pirivatra	kazivetar
kudibaba	mrzismet	COUNT= 0	letivetar
smrdibaba	probismet	čistikuca	COUNT= 2
visibaba	COUNT= 1	pazikuca	gulikoža
COUNT= 2		COUNT= 1	gulikoža
letivetar	palikuca	derikoža	COUNT= 1
COUNT= 0	pazikuca	gulikoža	jedibaba
luftiguz	COUNT= 1	COUNT= 1	smrdibaba
razbiguz	pazikuca	gazivoda	visibaba
trniguz	raspikuca	mutivoda	COUNT= 2
COUNT= 2	tecikuca	COUNT= 1	
mamipara	zatrikuca		
mamipara	COUNT= 3		

6. Conclusion

The results of the research presented in this paper are summarized in this section. Critical assessment of the comprehensive literature on the subject has shown that the problem of the structure of English and Serbo-Croatian noun compounds has not been treated in a satisfactory way. After a comprehensive study of theoretical works, reference works, grammar books and dictionaries, a corpus of 2527 English and Serbo-Croatian binomial hypotactic noun compounds has been selected to be subjected to analysis. The set of criterial distinctive features in the structure of the compounds has been hypothesized and the hypothesis tested by means of minimal pairs identified by the computer. The hypothesis has been systematically tested on the whole of the sample. An algorithm for the identification of minimal pairs has been written and implemented. The algorithm has been written in such a way that the hypothesis can be proven or rejected. The algorithm for the identification of minimal pairs in both languages has been proven to be successful. All hypothetical distinctive features in English have been shown (Table 2) and all but one in Serbo-Croat (Table 3). A further consequence of this has been that ambiguities (and even puns) can be treated in a computer-precise way (see, for example, minimal pairs: 'bogomoljka' (devout person) and 'bogomoljka' (an insect) in Table 9. and 'water-carrier' (person) and 'water-carrier' (instrument) in Table 6). The model proposed has been proven to be satisfactory; as a result it is plausible to suppose that it is a natural record of noun-compound reality. The results have further implications which are relevant in the fields of general linguistics, word-formation, machine translation, automatic speech recognition and language teaching.

REFERENCES

- Adams, V. 1973. *An introduction to Modern English word-formation*. London: Longman.
- Bauer, L. 1983. *English word-formation*. Cambridge: CUP.
- Belič, A. 1949. *Savremeni srpskohrvatski jezik 2. Nauka o gradjenju reci*. Beograd: Naučna knjiga.
- Bierwisch, M. and Heidolph, K. E. (eds). 1970. *Progress in linguistics*. The Hague: Mouton.
- Brekke, H. E. 1970. *Generative Satzsemantik und transformationelle Syntax im System der englischen Nominalkomposition*. Munchen: Wilhelm Fink Verlag.
- Charniak, E. and Wilks, Y. (eds). 1976. *Computational semantics: An introduction to artificial intelligence and natural language comprehension*. Amsterdam: North-Holland Publishing Company.
- Grishman, R. 1986. *Computational linguistics. An introduction*. Cambridge: CUP.
- Hansen, B. et al. 1982. *Englische Lexikologie. Einfuhrung in Wortbildung und lexikalische Semantik*. Leipzig: Veb Verlag Enzyklopadie.
- Janković, J. 1986. *Imenice slozenice sa glagolskom komponentom u engleskom i srpskohrvatskom jeziku*. Unpublished Ph.D. dissertation, Novi Sad.
- Janković, J. 1986. "The N-V dichotomy in the structure of English noun compounds?". *Literary and linguistic computing* 1/3. 143-155.
- Kelly, E. F. and Stone, P. J. 1975. *Computer recognition of English word senses*. Amsterdam: North-Holland Publishing Company.
- Leech, G. 1981. *Semantics*. 2nd ed. Harmondsworth: Penguin.
- Lees, R. B. 1970. "Problems in the grammatical analysis of English nominal compounds". In Bierwisch, M. and Heidolph, K. E. (eds). 1970. 174-186.

Stevanovič, M. 1964. *Savremeni srpskohrvatski jezik (gramatički sistemi i književnojezička norma)* 1. Beograd: Naučno delo.

Warren, B. 1978. *Semantic patterns of noun-noun compounds*. Goteborg: Gothenburg Studies in English 41.