

BESIM KABASHI
(Erlangen)

EINIGE NOTIZEN ÜBER DEN ROMAN *RRATHË* VON MARTIN CAMAJ

Martin Camaj ist ohne Zweifel ein Autor *sui generis*: Nicht nur aufgrund seines Lebenslaufes, sondern auch bezüglich seines wissenschaftlichen und künstlerischen Werkes. Er wählte stets einen eigenen Weg: Einen, der ihn berühmt machte und zugleich von anderen unterscheidet. Seine Werke sind besonders, sie sind durch ihre Originalität gekennzeichnet: Der Wortschatz, die Satzstruktur, der Stil, das Thema, und vieles andere.

Diese Eigenschaften waren die Motivation dafür, einige Statistiken zu einem seiner Werke, dem Roman *Rrathë* (dt. *Ringe*), vorzustellen, um dem Leser diese Aspekte von *Camajs* Werk, zumindest teilweise, bekannt zu machen.

Rrathë wurde 1978 in München im Selbstverlag publiziert. Der Roman ist in die drei Teile *Rrathë Uji* (dt. *Wasserringe*), *Rrathë Zjarmi* (dt. *Feuerringe*) und *Rrathë Gjaku* (dt. *Blutringe*) gegliedert, welche zusammen eine Einheit bilden.

Aufbereitung des Textes

Der Text¹ des Romans wurde maschinell verarbeitet. Zunächst wurde er zum Teil in herkömmlicher Form durchgeschaut und aufbereitet² und folglich in Sätze und in Wortformen zerlegt.³ Auf dieser Basis wurden die maschinellen Verfahren zur Analyse des Textes eingesetzt.

Wortformenebene

Der Text hat insgesamt 145488 laufende Textwörter. Der Anfang der Frequenzliste der Textwörter von *Rrathë* ist im Folgenden angegeben:

¹ Ich bedanke mich bei Herrn Prof. Dr. Bardhyl Demiraj (Universität München), der mir eine elektronische Version des Romans *Rrathë* zur Verfügung stellte, für wertvolle Gespräche bei der Ausarbeitung dieses Beitrages. Ebenso danke ich für wertvolle Gespräche Herrn Prof. Dr. Wilfried Fiedler (Meißen; Universität Jena).

² Dabei wurden einige nicht leichte Entscheidungen getroffen, z. B. wie bei einigen Dialogen die Satzgrenze festzustellen ist. Diese machen jedoch einen geringeren Teil des Gesamttextes aus, so dass die Statistiken zwar von der genannten Entscheidung beeinflusst sein könnten, jedoch nicht stark.

³ Die Analysen wurden größten Teils maschinell erstellt, weswegen die Ergebnisse als relativ anzusehen sind, da Fehler nicht ausgeschlossen werden können.

| | | | | |
|-----------|------------|------------|----------------|---------------|
| 8429 e | 8037 të | 4378 në | 2780 me | 2711 i |
| 1796 se | 1765 për | 1700 si | 1639 dhe | 1556 një |
| 1528 që | 984 prej | 921 u | 791 së | 777 tue |
| 702 nuk | 696 edhe | 657 ma | 637 pa | 553 kishte |
| 473 mbas | 461 ta | 461 po | 461 atë | 451 ishte |
| 426 do | 426 ai | 405 mbi | 390 tha | 384 t'i |
| 351 kish | 332 asht | 312 këtë | 311 ndër | 304 ia |
| 303 vetëm | 303 Agoni | 300 por | 291 apo | 289 Novelisti |
| 276 sa | 275 tij | 270 mos | 269 kur | 248 ajo |
| 247 herë | 237 nga | 237 më | 235 ka | 233 mirë |
| 232 mund | 231 deri | 228 dy | 220 sepse | 214 thanë |
| 210 ku | 208 kishin | 207 çdo | 203 Në | 201 iu |
| 195 vet | 188 E | 187 vete | 184 përpara | 182 qenë |
| 182 jo | 180 tjetër | 178 tyne | 178 mes | 177 ata |
| 171 para | 163 nëpër | 163 aq | 162 punë | 161 thoshte |
| 160 sy | 160 gjithë | 159 Por | 159 atje | 158 Si |
| 157 ditë | 152 as | 151 ky | 151 Edhe | 150 Dhe |
| 149 ashtu | 146 I | 145 ra | 143 vend | 143 Baci |
| 140 çka | 139 mori | 136 t'u | 133Folkloristi | 133 fjalë |
| 131 U | 128 kujtue | 128 kështu | 128 gjatë | 127 vesh |
| 127 Ripës | 126 ndoj | 125 marrë | 124 paska | 124 Bardhi |
| 123 saj | 121 unë | 121 nën | 120 diçka | 119 pasë |
| 119 kah | 118 ishin | 117 tepër | 117 kjo | 116 ato |
| 115 shumë | 115 aty | 113 tash | 111 vjetër | 111 sytë |
| 110 këtu | 110 ba | 109 vu | 108 kohë | 108 gjithnjë |
| 108 asaj | 107 dalë | 106 na | 106 duhet | 103 te |
| 102 atij | 101 sikur | 101 gja | 100 Mandej | 100 mandej |
| 100 kam | [...] | | | |

Liste 1 : Die häufigsten Wortformen aus Rrathë.

Diese Liste⁴ ist insgesamt 17039 Textwörter (Wortformen/Einträge) lang, wobei 9620 Wortformen nur einmal vorkommen (*hapax legomena*).⁵ Der Rest der Liste ist interessanter als der hier gezeigte Anfang, da die häufigsten Wörter meist Funktionswörter sind.

⁴ Diese Liste wird hier aus Platzgründen in gekürzter Form angegeben. (Für diese Zwecke ist eine elektronische Variante üblich).

⁵ Als Wortform wird eine Sequenz von Graphemen oder ein einziges Graphem getrennt durch ein „Leerzeichen“ definiert: Die Wortformen mit Apostroph, wie z. B. *t'i* werden als eine Wortform gezählt. Zwischen Klein- und Großbuchstaben wird bei dieser Zählung unterschieden, vgl. in der Liste 1 die Wortformen *Mandej* (100) und *mandej* (100).

Nach ihre Länge gruppiert ist die Verteilung der Types (Wortformen) und Tokens (alle Stellen einer Wortform im Text) in dem folgenden Diagramm und in der darauffolgenden Tabelle dargestellt:

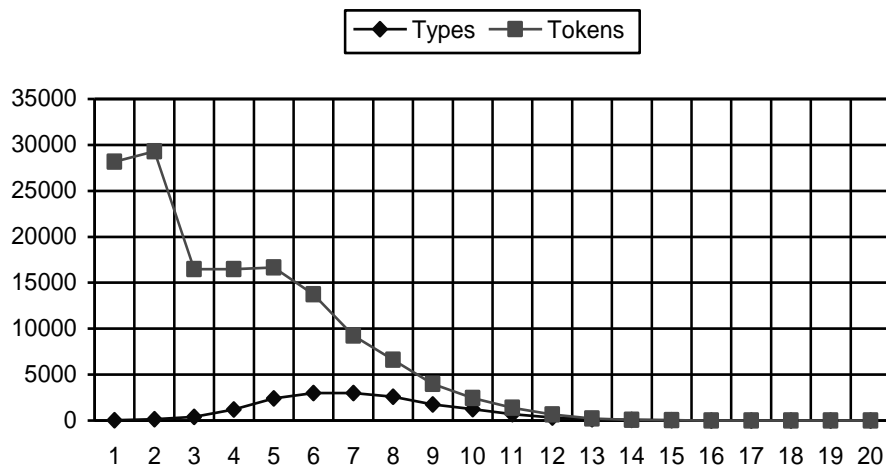


Abbildung 1 : Verteilung der Types und Tokens im Text anhand ihrer Länge.

Die Liste mit den Wortformen der Länge sieben ist die längste. Sie beinhaltet 3003 Types (Wortformen) mit 9239 Tokens (Stellen im Text). Die ersten fünf sind *përpara* (mit 184 Vorkommen im Text), *thoshte* (161), *vërtetë* (62), *diellit* (62) und *ndëgjue* (59). Die längste Wortform ist *shtatëdhjetëvjeçarit* (mit 20 Graphemen; [18 Buchstaben]), welche nur einmal im Text vorkommt. Ebenso die zweitlängste Wortform (19), die zum selben Wortparadigma gehört: *shtatëdhjetëvjeçarë*. Die drittlängste Gruppe umfasst ebenso nur eine Wortform, nämlich *politiko-ekonomike* (18), welche auch nur einmal im Text vorkommt. Die Gruppe mit der Länge 17 hat sechs Types mit insgesamt sieben Tokens, wobei die Wortform *bashkëpunëtorëvet* zweimal vorkommt. Die anderen fünf Tokens sind *pendëzhdërvjellët*, *paqëndrueshmëninë*, *mbrendapërmbrenda*, *shtatëdhjetëvjeçar* und *përgjithësoheshin* mit je einem Vorkommen. Die nächste kleinere Gruppe, Länge 16, enthält fünf Types mit sechs Tokens. Der Type *bashkëpunëtorëve* kommt zweimal vor, die restlichen vier *drejtpërdrejtazi*, *mirë-tingëlluese*, *tetëdhjetëvjeçare* und *paqëndrueshmënia* je einmal. Die Zahl der Einträge in anderen Gruppen, vgl. hierzu die Tabelle 1, ist deutlich größer. Die durchschnittliche Wortlänge ist 3.94339, mit einer Standardabweichung von 2.59393.

| Länge, Types, Token | | | | |
|---------------------|-----------------|-----------------|------------------|------------------|
| 1 / 36* / 28184 | 2 / 143 / 29284 | 3 / 389 / 16459 | 4 / 1219 / 16481 | 5 / 2396 / 16646 |
| 6 / 2979 / 3746 | 7 / 3003 / 9239 | 8 / 2576 / 6604 | 9 / 1768 / 3993 | 10 / 1251 / 2465 |
| 11 / 694 / 1387 | 12 / 340 / 642 | 13 / 148 / 206 | 14 / 62 / 86 | 15 / 30 / 50 |
| 16 / 5 / 6 | 17 / 6 / 7 | 18 / 1 / 1 | 19 / 1 / 1 | 20 / 1 / 1 |

Tabelle 1 : Verteilung der Types und Tokens im Text anhand ihrer Länge.⁶

Es wurden weitere statistische Berechnungen⁷ über die Sprache von *Rrathë* erstellt, wie *Kookkurrenzen* auf Satzebene unter Verwendung der Assoziationsmaße *Log-Likelihood*, *Chi-Quadrat* und *Observed Expected*, welche hier aus Platzgründen nicht gezeigt werden können.⁸ Ein Ausschnitt aus der einfachen *Kookkurrenz-Analyse*, sortiert nach ihrer Häufigkeit, ist als Anhang angegeben.

Außer den angegebenen Statistiken im Bereich des Lexikons bzw. der Morphologie, d. h. auf der Wort(form)ebene, fällt, schon bei den wenigen gezeigten Einträgen auf, dass *Camaj* eine eigene Rechtschreibung benutzt, eine Orthographie, die sorgfältig erstellt wurde und eine Normierung anstrebt, vgl. hierzu die in Kursivschrift hervorgehobenen Einträge in der Liste 1. *Camajs* Bestrebungen, nicht nur im Falle von *Rrathë*, sondern auch bei anderen literarischen und bei linguistischen Werken, wie der *Albanian Grammar*, gelten dem Norddialekt des Albanischen, dem *Gegischen*.⁹ Dies lässt sich ausser in den Einträgen in der Liste 1 auch in folgenden, aus der Wortformliste von *Rrathë* ausgeschnittenen, Beispielen erkennen: *hangër*, *hangërkan*, *hangërme*, *hangra*, *hangri*, *hangri*, *hangun*, *hanguna* und *hangunit*.

Bei der Erhebung der Statistiken fielen bezüglich der Orthographie u. a. auch die Wortformen *mb'anë* (9×, Seiten 194, 224 (2×), 226 (2×), 235, 263, 266 und 268), *mb'anësh* (3×, Seiten 259, 261 und 297), *mb'atë* (1×, Seite 269) und *q'andëj* (8×, Seiten 16, 59, 63, 100, 140, 193, 243 und 275) auf.

Ein anderer Aspekt auf der Wortebene ist die Wortbildung. Den Wortformen *dhelpën*, *dhelpnake*, *dhelpnash*, *dhelpne*, *dhelpnia*, *dhelpnie*, *dhelpninë*, *dhelpnisht* liegt das gleiche Lexem zugrunde, welches eine reiche Wortbildung zeigt. Als weiteres Beispiel könnten die folgenden Wort-

⁶ Bei dieser Zählung (Länge 1, *) wurden auch die Interpunktionszeichen mitgezählt.

⁷ Eine Erklärung dieser und weiterer Fachbegriffe aus diesem Fachbereich findet sich z. B. bei [MANNING/SCHÜTZE 1999], [EVERT ET AL. 2010] und [HEYER ET AL. 2008].

⁸ Für wertvolle technische Hinweise bei der Erstellung der Statistiken danke ich meinen Kollegen Thomas Proisl und Peter Uhrig (Universität Erlangen-Nürnberg).

⁹ Martin Camaj : *Albanian Grammar*. Wiesbaden : Harrassowitz, 1984.

formen dienen: *bashkë, bashkëfetarët, bashkëfjalim, bashkëfolësit, bashkëkohore, bashkënxanës, bashkëpunëtor, bashkëpunim, bashkëpunojë, bashkërenditëse, bashkëshortëve, bashkësi, bashkëvendës, bashkëveprues, bashkim, bashkohen, bashku, bashkue* und *bashkueme*.¹⁰

Ein besonderes Merkmal bei *Camaj* sind Wortformen bzw. Wörter i. S. v. Lexemen, die selten vorkommen wie z. B. *cilikoka* (Seiten 122 und 133 [*cilikokë*]) und in die Norm-Sprache, etwa in [FJALOR 1980], nicht aufgenommen sind – auch nicht als regional-dialektale Lemmata/Einträge.

Satzebene

Der Text beinhaltet insgesamt 6343 Sätze mit einer durchschnittlichen Länge von 22.66562 und einer Standardabweichung von 13.91838. Der längste Satz hat – unter der Einbeziehung der Satzzeichen – 145 Wortformen.¹¹

145: *Ky krahasim iu ndër mend Bardhit të cilit i dhimbtën sytë vetëm tue shikue prej asaj naltësie njerinë që s'lidhej as për për brinash , as për fjale , shfrytëzues i çdo sistemi , buntar i pa pagesë në pëllasin ma të shkëlqyeshëm e në tëbanin ma të lanun , ngranës kavjari dhe henash të zgjedhuna ndër tryeza të shtrueme për luks , me thika e piruj argjandi , por edhe djathi me shtrepa e bukë misri të rreshkun pranë zjarmit ndër votrat malësore ku kalonte një pjesë të vjetit tue mbledhë kangë trimnish , përralla e vajtime grash ; përdorës ndër udhëtime të veta të makinës , të karrocës e të qerres me dy rrota të mëdha që tërhiqet prej qeve , apo i gomarit të vjetër e të pafuqishëm të cilit me peshën e trupit të vet ia palonte shpinën pa farë dhimbe . (Rrathë, S. 32)*

Der zweitlängste Satz ist der folgende:

118: *Si pa u kujtue gëzimi i shkëputjes prej asaj bote ndeshi në një shije irnuese në thelbin e zemrës së Novelistit dhe nga këto parandjenja nisën të shkëputen tallaze dyshimesh , ndër të tjera se si ai përpiqesh më kot t'i largohej qytetit si anijë në midis valëve të detit dhe se ai mbas shumë kohe , në vendbanimin mbas malit , në thellësinë e gurrave malore , do ta ndiente veten si njeri i gjymtuem , për shembull , pa kambë , jo ma i zoti për t'u kapë përsëri shpatit të malit prej asaj ane deri aty ku ishte tash dhe q'andej për ta shijue pamjen e qytetit , sa për të hjekë pak mall . (Rrathë, S. 275)*

Von Interesse wären bei diesen Sätzen u. a. auch deren Strukturen, die Sequenzen sowie die Abhängigkeiten der Satzglieder von einander und von zugehörigen Verben.

Der nächste Satz (mit überdurchschnittlicher Länge), zufällig ausge-

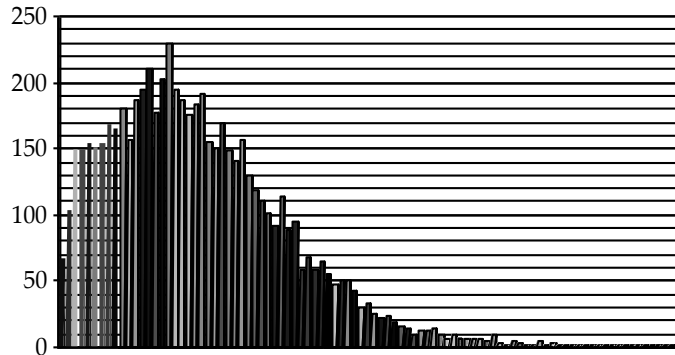
¹⁰ Hier wurden nur einige der Derivata und Komposita von *bashkë* angegeben.

¹¹ Der Satz enthält höchstwahrscheinlich einen Druckfehler, vgl. die Hervorhebung der Wortform *për*.

wählt, macht z. B. außer dem Punkt als Satz-Ende-Markierung, keinen anderen Gebrauch von Interpunktionen:

Në kashtën e kumbarës ishin ngjitë si kokrra në shëgën e çame tufa yjësish dhe ashtu të kapun dendas shndritnin aq sa kishte thanë se qielli qenkej shkye dikund në një vend e prej andej del dritë si prej një dritarje në katund . (Rrathë, S. 86).

Die Verteilung der Sätze im Text anhand ihrer Länge sieht wie folgt aus:



| | | | | | | | | | | | |
|----|----|----|-----|-----|-----|-----|-----|-----|-----|----|----|
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 |
| 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 |
| 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 |
| 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 |
| 74 | 75 | 76 | 77 | 78 | 80 | 81 | 82 | 83 | 84 | 86 | 87 |
| 89 | 96 | 99 | 100 | 102 | 103 | 105 | 114 | 118 | 145 | | |

Abbildung 2 : Verteilung der Sätze im Text anhand ihrer Länge.

Die Sätze der Länge 18 stellen mit 230 Sätzen die größte Gruppe dar. Die zweitgrößte Gruppe bilden 203 Sätze mit der Länge 17. Die genaueren Daten der Satzlänge sind in der folgenden Tabelle angegeben:

| Länge / Sätze | | | | | | | |
|---------------|----------|----------|----------|----------|----------|----------|----------|
| 2 / 66 | 3 / 103 | 4 / 149 | 5 / 149 | 6 / 153 | 7 / 151 | 8 / 153 | 9 / 168 |
| 10 / 165 | 11 / 181 | 12 / 156 | 13 / 186 | 14 / 194 | 15 / 211 | 16 / 177 | 17 / 203 |
| 18 / 230 | 19 / 195 | 20 / 186 | 21 / 175 | 22 / 183 | 23 / 192 | 24 / 155 | 25 / 151 |
| 26 / 170 | 27 / 148 | 28 / 141 | 29 / 157 | 30 / 129 | 31 / 118 | 32 / 111 | 33 / 102 |
| 34 / 92 | 35 / 114 | 36 / 89 | 37 / 95 | 38 / 59 | 39 / 68 | 40 / 59 | 41 / 65 |
| 42 / 55 | 43 / 48 | 44 / 51 | 45 / 51 | 46 / 42 | 47 / 30 | 48 / 33 | 49 / 26 |
| 50 / 22 | 51 / 24 | 52 / 19 | 53 / 16 | 54 / 15 | 55 / 10 | 56 / 12 | 57 / 12 |

| | | | | | | | |
|---------|---------|---------|---------|---------|---------|--------|---------|
| 58 / 15 | 59 / 9 | 60 / 6 | 61 / 9 | 62 / 7 | 63 / 7 | 64 / 6 | 65 / 7 |
| 66 / 5 | 67 / 9 | 68 / 3 | 69 / 2 | 70 / 5 | 71 / 3 | 72 / 1 | 73 / 2 |
| 74 / 5 | 75 / 1 | 76 / 3 | 77 / 2 | 78 / 1 | 80 / 2 | 81 / 1 | 82 / 1 |
| 83 / 1 | 84 / 1 | 86 / 1 | 87 / 2 | 89 / 2 | 96 / 1 | 99 / 1 | 100 / 1 |
| 102 / 1 | 103 / 1 | 105 / 1 | 114 / 1 | 118 / 1 | 145 / 1 | | |

Tabelle 2 : Verteilung der der Sätze im Text anhand ihrer Länge.

Die kurzen Sätze sind oft für linguistische Forschungszwecke interessant. Im Folgenden sind Ausschnitte aus den Sätzen der Länge zwei und drei angegeben:

| | | | |
|--------|------------|--------------|-----------|
| Besa ! | Lene ! | Njelmësi ! | Qite ! |
| Fol ! | Merr ! | Përnjimend ! | Rrehash ! |
| Hej ! | Mjaft ! | Pije ! | Si ? |
| Heu ! | Mue ? | Pikërisht ! | Shiko ! |
| Hi ? | Ndal ! | Pyeti . | Vërtet ! |
| Jo ! | Ngarkuet ! | Qindro ! | Tmerr ! |

Liste 2 : Ausschnitt aus der Gruppe der Sätze mit Länge 2.

| | | |
|-----------------|--------------------|-------------------|
| Ajo asht ! | Ju jini ? | Po dihet ! |
| Apo dëshprimi ? | Kjo pra ! | Po ti ? |
| Asht kryetari . | Ktheu dalë . | Po ty ? |
| Ashtu asht ! | Kush asht ? | Priti gjatë . |
| Çfarë ka ? | Kush je ! | Pse jo ? |
| Cili njeri ? | Leji mrekullitë ! | Punë fëmijësh ! |
| Çka dëshironi ? | Leni dredhitë ! | Pyeti Agoni . |
| Çka flitet ? | Ma tepër ? | Pyetje retorike ! |
| Çka kemi ? | Mbi kë ? | Pyet ti . |
| Dhe pse ? | Më lësho ! | Qoftë larg ! |
| E fejuemja ? | Merr vesh ! | Shiko këtu ! |
| E mandej ? | Më të ! | Shumë larg ! |
| E nxuna ! | Na turpnove ! | Si autodidakt ? |
| Gja epike ! | Në pupla ? | Si kështu ? |
| Hiqe maskën ! | Ngjeshi armët ! | S'ka gja ! |
| Ide ndërtuese ? | Njëzet minuta ? | S'po kuptohemi ! |
| I jati ! | Pa turp ! | Të gjithë ? |
| Ishte verë . | Përfundoi Agoni . | Të kuptoj . |
| Iu përgëztoj ! | Përtej kufijve ? | Tha ai . |
| Jemi heret . | Pëshpëriti nusja . | Thanë ata . |

Thirri Drënashi .
Thoshte Sosja .
Thotë kjo !

U gëzue .
U ndanë .
U vesh .

Vërejti Novelisti .
Vetëm fjalë ?
Vetëm hi ?

Liste 3 : Ausschnitt aus der Gruppe der Sätze mit Länge 3.

Schlussbemerkungen

Obwohl hier die vollständige Analyse des Romantextes auf wenigen Druckseiten nicht untergebracht werden konnte, bleibt zu hoffen, dass sich der Leser aus den wenigen Stichproben und Ausschnitten einen ersten Eindruck verschaffen kann.

Um aus diesen und weiteren Statistiken gewünschte Schlüsse zu ziehen z. B. eine Differenzanalyse durchzuführen, wäre ein Referenzkorpus notwendig gewesen, das für das Albanische leider immer noch fehlt. Eine detaillierte automatische Wortformanalyse bzw. die Klassifizierung der Wortformen in Klassen bzw. ihre Zuordnung zu Paradigmen ist ohne ein automatisches Morphologie-System nicht möglich. Solche gibt es inzwischen für das Norm-Albanische, vgl. hierzu z. B. [KABASHI 2004] und [KABASHI 2007 Z], jedoch nicht für den gegischen Dialekt oder in diesem Fall für die Sprache *Camajs*.

Aus den Ergebnissen der verwendeten korpuslinguistischen Methoden bzw. aus den wenigen hier vorgestellten Daten könnte man zusammenfassend sagen, dass *Camaj* in erster Linie, aus linguistischer Sicht¹² betrachtet, eine Sprache „ausgewählt“ hat, welche ein starkes gegisches Element aufweist, jedoch, sowohl im Wortschatz als auch auf der Satzebene, eine Nähe zur eine Norm-Sprache anstrebt. Diese unternimmt einen Schritt auf ein vereinheitlichtes Albanisch hin – eine nicht leichte „Aufgabe“. Eine solche Sprache kann in einem Schriftmedium nicht besser als in der Form eines Romans wiedergegeben werden. Und, in dem Roman *Rrathë* hat *Martin Camaj*, die von ihm angestrebte Sprache für die zukünftigen Generationen festgehalten.

Referenzen und Literatur

[EVERT ET AL. 2010]

Evert, Stefan / Frötschl, Bernhard / Lindstrot, Wolf : „Statistische Grundlagen“. 114–158. [In:] Carstensen, Kai-Uwe / Ebert, Christian / Ebert, Cornelia / Jekat, Susanne / Klabunde, Ralf / Langer, Hagen (Hrsg) : *Computerlinguistik und Sprachtechnologie. Eine Einführung*. 3. Auflage. Heidelberg : Spektrum Akademischer Verlag, 2010.

¹² Der literarische Wert wird hier keineswegs diskutiert.

[FJALOR 1980]

Kostallari, Androkli et al. : *Fjalor i gjuhës së sotme shqipe*. Tiranë : Akademia e Shkencave e RPS të Shqipërisë, 1980.

[HEYER ET AL. 2008]

Heyer, Gerhard / Quasthoff, Uwe / Wittig, Thomas : *Text Mining : Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse*. Herdecke / Bochum : W3L-Verlag, 2008.

[KABASHI 2004]

Kabashi, Besim : „Analiza automatike e fjalëformave të gjuhës shqipe“. 129–135. [In:] *Seminari Ndërkombëtar për Gjuhën, Letërsinë dhe Kulturën Shqiptare, XXIII*. Prishtinë : Universiteti i Prishtinës, 2004. *Libri 23/1*.

[KABASHI 2007 Z]

Kabashi, Besim : „Zeichen für Gjon Buzuku. Die Zusammenarbeit zwischen der albanischen Linguistik und der Computerlinguistik.“ 139–146. [In:] Demiraj, Bardhyl (Hrsg.) : *Nach 450 Jahren. Buzukus „Missale“ und seine Rezeption in unserer Zeit*. (Albanische Forschungen 25). Wiesbaden : Harrasowitz, 2007.

[MANNING/SCHÜTZE 1999]

Manning, Christopher D. / Schütze, Hinrich : *Foundation of Statistical Natural Language Processing*. Cambridge, Massachusetts : MIT Press, 1999.

Anhang

| | | | | | |
|-----------|-----------|-------|------------|------------|------|
| të (8038) | të (8038) | 16004 | i (2711) | të (8038) | 4355 |
| të (8038) | e (8429) | 15678 | në (4378) | në (4378) | 3902 |
| e (8429) | të (8038) | 15678 | të (8038) | për (1765) | 3664 |
| e (8429) | e (8429) | 15388 | për (1765) | të (8038) | 3664 |
| në (4378) | e (8429) | 8687 | të (8038) | si (1700) | 3419 |
| e (8429) | në (4378) | 8687 | si (1700) | të (8038) | 3419 |
| të (8038) | në (4378) | 8136 | e (8429) | dhe (1639) | 3314 |
| në (4378) | të (8038) | 8136 | dhe (1639) | e (8429) | 3314 |
| të (8038) | me (2780) | 5433 | si (1700) | e (8429) | 3248 |
| me (2780) | të (8038) | 5433 | e (8429) | si (1700) | 3248 |
| me (2780) | e (8429) | 5181 | të (8038) | dhe (1639) | 3200 |
| e (8429) | me (2780) | 5181 | dhe (1639) | të (8038) | 3200 |
| i (2711) | e (8429) | 4402 | se (1796) | e (8429) | 3064 |
| e (8429) | i (2711) | 4402 | e (8429) | se (1796) | 3064 |
| të (8038) | i (2711) | 4355 | për (1765) | e (8429) | 3046 |

| | | |
|-----------------|---------------|---------|
| e (8429) | për (1765) | 3046 |
| të (8038) | një (1556) | 3044 |
| një (1556) | të (8038) | 3044 |
| që (1528) | e (8429) | 2999 |
| e (8429) | që (1528) | 2999 |
| të (8038) | që (1528) | 2990 |
| që (1528) | të (8038) | 2990 |
| të (8038) | se (1796) | 2963 |
| se (1796) | të (8038) | 2963 |
| një (1556) | e (8429) | 2689 |
| [...] | [...] | [...] |
| ujit (32) | e (8429) | 100 |
| tepër (117) | ma (657) | 100 |
| të (8038) | sistemit (44) | 100 |
| të (8038) | krah (66) | 100 |
| sistemit (44) | të (8038) | 100 |
| sikur (101) | në (4378) | 100 |
| si (1700) | qenë (182) | 100 |
| se (1796) | ata (177) | 100 |
| qenë (182) | si (1700) | 100 |
| por (300) | një (1556) | 100 |
| për (1765) | mirë (233) | 100 |
| para (171) | me (2780) | 100 |
| një (1556) | por (300) | 100 |
| në (4378) | sikur (101) | 100 |
| në (4378) | kah (119) | 100 |
| mirë (233) | për (1765) | 100 |
| mes (178) | i (2711) | 100 |
| me (2780) | para (171) | 100 |
| me (2780) | ky (151) | 100 |
| ma (657) | tepër (117) | 100 |
| ky (151) | me (2780) | 100 |
| krah (66) | të (8038) | 100 |
| kah (119) | në (4378) | 100 |
| i (2711) | mes (178) | 100 |
| e (8429) | ujit (32) | 100 |
| ata (177) | se (1796) | 100 |
| [...] | [...] | [...] |
| sjellë (44) | të (8038) | 50 |
| si (1700) | Por (159) | 50 |
| si (1700) | kështu (128) | 50 |
| Si (158) | një (1556) | 50 |
| shënueshme (28) | e (8429) | 50 |

| | | |
|----------------|---------------|---------|
| së (791) | Poshtme (51) | 50 |
| qiell (26) | të (8038) | 50 |
| qenka (48) | i (2711) | 50 |
| qenë (182) | që (1528) | 50 |
| që (1528) | qenë (182) | 50 |
| që (1528) | jo (182) | 50 |
| prandej (28) | të (8038) | 50 |
| Poshtme (51) | së (791) | 50 |
| por (300) | ai (426) | 50 |
| Por (159) | si (1700) | 50 |
| pikë (35) | në (4378) | 50 |
| përpiqej (25) | të (8038) | 50 |
| paskej (33) | të (8038) | 50 |
| parën (27) | e (8429) | 50 |
| njomë (13) | të (8038) | 50 |
| njeri (82) | me (2780) | 50 |
| njerëzish (24) | të (8038) | 50 |
| një (1556) | Si (158) | 50 |
| një (1556) | anë (53) | 50 |
| [...] | [...] | [...] |
| lumtun (11) | i (2711) | 10 |
| lumit (47) | ndër (311) | 10 |
| lumit (47) | andej (99) | 10 |
| lumë (14) | i (2711) | 10 |
| lulonin (1) | e (8429) | 10 |
| lulesh (1) | e (8429) | 10 |
| luginë (7) | në (4378) | 10 |
| lugina (3) | e (8429) | 10 |
| lugë (3) | të (8038) | 10 |
| [...] | [...] | [...] |
| arsye (30) | kisha (24) | 1 |
| arsye (30) | këto (73) | 1 |
| arsye (30) | ketë (27) | 1 |
| arsye (30) | kështu (128) | 1 |
| arsye (30) | këshillue (4) | 1 |
| arsye (30) | kërkonte (20) | 1 |
| [...] | [...] | [...] |