

# Palomar: a Computer-aided Analysis of some Lexical and Stylistic Features

Adriana Roventini

---

Istituto di Linguistica Computazionale – ILC –  
Pisa, Via della Faggiola 32, Pisa, Italy

KEYWORDS: computer-aided stylistic analysis

E-MAIL: adriana@ilc.pi.cnr.it  
FAX NUMBER: +39 50 589055  
PHONE NUMBER: +39 50 560481

## Abstract

This paper presents a few initial results of a study which is being carried out at the Institute of Computational Linguistics in Pisa, fitting into the framework of computer-based approaches to literary text analysis. The object of this investigation is Palomar, a work of the well-known Italian writer Italo Calvino. This book, published in 1983, is one of Calvino's later works and is very representative of the author's stylistic and thematic aspects. The paper focuses on the lexical structure and the particular use of the different categories on the syntagmatic level (i.e. the way they combine), as well as the computer tools used to exploit this information in a consistent and exhaustive way.

## 1. The Pisa tools for text analysis

In the last few years, at ILC in Pisa, a set of tools, known as the PiSystem, has been developed and continuously improved in order to make possible the exploitation of text corpora according to the most varied user needs. In particular, a statistically based procedure has been implemented for the automatic lemmatization and tagging of Italian texts: the Pitagger (Picchi 1994). This procedure assigns the grammatical category and the lemma to the word forms occurring in whatever Italian text is being analyzed. Obviously, the output of this fully automatic procedure is not completely correct. Furthermore, if a deep morphological analysis is requested, a choice must be made among, for example, the homograph inflected verb forms. So, in order to obtain a correct and sufficiently complete text analysis, as in the case of studies aiming to individuate and describe stylistic features, an interactive procedure, called Tagghand, helps the user to correct and to deepen his/her research. After that, another interactive procedure, the Dblemcor, allows the user to make new changes on the already lemmatized word forms. This is a very useful tool because this kind of work is subject to doubts and afterthoughts. The difficulty

of choice between different grammatical categories when dealing with borderline cases such as past participle / adjective, infinitive / noun, present participle / adjective, present participle / noun and so on is well known by anyone who has ever done a lemmatization.

In the work presented here both procedures (the Tagghand and the Dblemcor) were used and a fine-grained coding was performed to allow further and more sophisticated researches relating to grammatical categories, their employ and their sequential combinations, from a stylistic point of view. To this end, procedures offering the possibility of selecting the lemmata according to either grammatical categories or morphological codes, besides the various ending strings, turned out to be very helpful. In the annexed figures a few examples of different subsets of this lexicon are shown. They were obtained by querying for:

- the feminine singular nouns derived by the suffix *-ezza*, Fig. 1;
- the masculine singular adjectives derived by the suffix *-ico*, Fig. 2;
- the inflected verbal forms carrying as morphologic code "S3IP", Fig. 3;
- the word forms carrying as grammatical code "W", i.e. the proper nouns, Fig. 4.

## 2 Lexical features

According to unanimous critical judgement, Calvino's writing is characterized by great clarity, accuracy and punctuality, which should be reflected in his lexical choices (see Mengaldo 1991 and 1994). So, in the survey on Palomar's lexical structure, the use of the computer appeared particularly useful to verify and quantify, when possible, critical intuitions and suggestions. In the limited space of this presentation only a few features will be considered, in particular:

- the creation of new words to outline particular ideas, minds or sensations;
- the preference accorded to certain types of word formations, and, vice-versa, the limited use of others (e.g. the diminutives).

As far as the first feature is concerned, the data extracted from the text confirms many critical remarks asserting that, normally, the creativity of this author is not expressed by means of new words, but rather relies on an exceptional linguistic competence (also as far as scientific sublanguages are concerned) joined with a particular ability to stress all the possibilities of those syntactic combinations which the Italian language plentifully allows for. This seems to be true also for Palomar. In fact, the new words, which constitute a little subset within this lexicon, are mostly original matchings of common words by hyphenation, and,

if we do not consider those adjectives expressing different color shades, such as *bronzео-roseа* (bronze-rosy) or *nero-giallastro* (black-yellowish), we have fourteen new compounds in all, aiming to express particular thoughts, opinions and feelings of the writer, listed and literally translated below:

<i>braccia-zampe</i>	(arms-paws)
<i>continenti-cimiteri</i>	(continents-cemeteries)
<i>divergenza-convergenza</i>	(divergence-convergence)
<i>finestra-vetrina</i>	(window-shop window)
<i>gabbia-vetrina</i>	(cage-shop window)
<i>lumpen-pennuti</i>	(lumpen-feathered)
<i>spiaggia-immondezzaio</i>	(plage-garbage)
<i>silenzio-discorso</i>	(silence-speech)
<i>umanità-sabbia</i>	(humanity-sand)
<i>uomo-albero</i>	(man-tree)
<i>uomo-bue</i>	(man-ox)
<i>iperlavoro</i>	(hyperwork)
<i>ultralavoro</i>	(ultrawork)

Regarding the second feature, the lists of nouns and adjectives (extracted according to their inverse alphabetical order) prove that diminutives, augmentatives and other types of alterations, belonging to an emotional more than a precise and punctual language, are scarcely used. In the nouns list only the following altered word-forms appear:

<i>farfallina</i>	(little butterfly)
<i>macchiolina</i>	(little spot)
<i>fogliolina</i>	(little leaf)
<i>pianticelle</i>	(little plants)
<i>cascatelle</i>	(little waterfalls)
<i>lombatine</i>	(little sirloins)
<i>scimmione</i>	(big monkey)
<i>fiorellini</i>	(little/nice flowers)
<i>pilastrini</i>	(little pilasters)
<i>ragazzotti</i>	(big/rough boys)
<i>praticello</i>	(little/nice lawn)
<i>bastoncino</i>	(little stick)

The adjectives subset shows a few superlatives and four alterates in *-ino* and *-astro*:

<i>velocissima</i>	(very quick)
<i>elaboratissima</i>	(very elaborate)
<i>rapidissime</i>	(very rapid)
<i>bellissime</i>	(very beautiful)
<i>vicinissimi</i>	(very near)
<i>nitidissimo</i>	(very clear)
<i>bianchissimo</i>	(very white)
<i>difficilissimo</i>	(very difficult)
<i>minutissimo</i>	(very minute)
<i>biancastra</i>	(whitish)
<i>giallastro</i>	(yellowish)
<i>freddina</i>	(chilly)
<i>giallini</i>	(faded yellow)

The possibility of selecting subsets of lemmatized word forms in inverse alphabetical order, according to either their syntactic categories or to a particular derivative suffix, evidenced a great variety and richness in the use of derived nouns and adjectives which appear to be, on the lexical level, Calvino's main way of obtaining accuracy and

exactitude.

Because it is impossible here to draw up complete lists of them, we will provide some quantitative data and we will comment on only a few of the more interesting and widely used examples. The analysis of the complete list of the inverse-ordered nouns shows that the most frequent derivative suffixes attested within this lexicon are the following:

<i>-sione(i)/-zione(i)</i>	205
<i>-ità/età/tà</i>	80
<i>-anza(e)/-enza(e)</i>	79
<i>-mento(i)</i>	60
<i>-ezza(e)</i>	36
<i>-sura(e)/-tura(e)</i>	31

A first division of these suffixes can be made and we can state that three of them (*-zione*, *-mento*, *-tura*) are deverbal and derive actions or processes nouns. The other three suffixes (*-ità*, *-anza*, *-ezza*) are deadjectival and give rise to nouns expressing qualities and conditions. In looking at these sets of derived nouns, we can confirm a typical character attributed to Calvino's writing: many words belong to scientific sublanguages, testifying to his strong interest in the natural, physical, mathematical and anthropological sciences. These terms abound, especially within the first group of deverbal nouns while, in the second group, scientific terms are less frequent, even if a few terms typical of physical and chemical fields were found, as can be seen from the lists below:

<i>corrosione</i>	(corrosion)
<i>dispersione</i>	(dispersion)
<i>dissociazione</i>	(dissociation)
<i>disgregazione</i>	(disintegration)
<i>assimilazione</i>	(assimilation)
<i>inclinazione</i>	(inclination)
<i>migrazione</i>	(migration)
<i>pigmentazione</i>	(pigmentation)
<i>distribuzione</i>	(distribution)
<i>classificazione</i>	(classification)
<i>variazione</i>	(variation)
<i>angolazione</i>	(angle)
<i>proliferazione</i>	(proliferation)
<i>segmentazione</i>	(segmentation)
<i>intersezione</i>	(intersection)
<i>induzione</i>	(induction)
<i>condizionamento</i>	(conditioning)
<i>comportamento</i>	(behaviour)
<i>procedimento</i>	(process)
<i>mutamento</i>	(change)
<i>ragionamento</i>	(reasoning)
<i>esperimento</i>	(experiment)
<i>punteggiatura</i>	(punctuation)
<i>nomenclatura</i>	(nomenclature)
<i>temperatura</i>	(temperature)
<i>partitura</i>	(score)
<i>scrittura</i>	(writing)
<i>struttura</i>	(structure)
<i>solidità</i>	(solidity)
<i>fluidità</i>	(fluidity)
<i>potenzialità</i>	(potentiality)
<i>granulosità</i>	(granularity)
<i>continuità</i>	(continuity)
<i>umidità</i>	(humidity)
<i>profondità</i>	(depth)

densità	(density)
luminosità	(luminosity)
discontinuità	(discontinuity)
discrepanza	(discrepancy)
tolleranza	(tolerance)
sequenza	(sequence)
trasparenza	(transparency)
esperienza	(experience)
consistenza	(consistency)
lunghezza	(length)
altezza	(height)
completezza	(completeness)
ampiezza	(amplitude)
lucentezza	(brilliance)
compattezza	(compactness)

Another interesting point evidenced by the simple survey on the lexicon composition is that some scarcely productive types of suffixes can also be relevant in the research of the links between lexicon and style because they are carriers of typical features. This is the case with the deadjectival suffix *-ore(i)* and the denominal suffix *-io(ii)*. The first is employed in the semantic field of colours and visual perception, the second carries a particular aspectual and intensive value in both the fields of vision and audition. So they are clues of a central theme in this work: the mechanisms of perception (in particular the sight) as source of knowledge; Palomar being an attentive, careful and nearly maniacal observer of the world surrounding him. Besides, they confirm a critical suggestion by Mengaldo about a certain preference for the derivatives in *-io*. Below are a few examples literally (when possible) translated:

biancore	(whiteness)	brontolio	(grumbling)
bagliore	(flare)	cinguettio	(twittering)
splendore	(splendour)	borbottio	(mumbling)
fulgore	(splendour)	chioccolio	(gurgling)
grigiore	(greyness)	tremolio	(twinkling)
chiarore	(glimmer)	scintillio	(glitter)

Obviously, the list of the word frequencies is the first way to individuate thematic features and, in the Palomar frequencies list, we find that *sguardo* (look) is, after *mondo* (world) and *cielo* (sky), the third more frequent semantic word. Nevertheless, it is interesting to detect similar indications also at different linguistic levels, such as the derivatives word classes.

Remarks and comments of the same kind can be done for the adjectives. They also present many derivatives typical of scientific languages (*-bile*, *-ico*) and some others which are more literary or latinizing (see *-eo* for example). The deverbals (*-bile*, *-ente*, *-ivo*), and the denominals (*-ico*, *-oso*, *-eo*) were among the most frequent suffixes found. Below quantitative indications and a few examples are given:

-ico(a,he,i)	100
-bile(i)	66
-ente(i)	57
-oso(a,e,i)	52
-ivo(a,e,i)	34
-eo(a,e,i)	31

probabile	(probable)	variabile	(variable)
tangibile	(tangible)	visibile	(visible)
sensibile	(sensible)	osservabile	(observable)
cosmico	(cosmic)	biologico	(biological)
allegorico	(allegorical)	simmetrico	(symmetric)
geometrico	(geometric)	simbolico	(symbolic)
erboso	(grassy)	granuloso	(granulose)
cremoso	(creamy)	poroso	(porous)
gustativo	(gustative)	olfattivo	(olfactory)
uditivo	(auditive)	visivo	(visual)
erbaceo	(herbaceous)	bronzeo	(bronze)
aereo	(airy)	argenteo	(silvery)
vitreo	(vitreous)	corporeo	(bodily / corporeal)
roseo	(rosy)	omogeneo	(homogeneous)
latteo	(milky)	sidereo	(sidereal)

### 3. The way Calvino combines the words: the adjectivation

As stated above, the other principal way used by Calvino to reach extraordinary accuracy and flexibility in his writing is through the use of a ductile and varied combination of grammatical categories within the sentences.

The Italian language allows, in this sense, great possibility of variation, and the adjectivation in Palomar is a perfect example of this. A procedure which allows the retrieval of grammatical combinations evidenced many schemes of noun-adjective association. Below the more frequent patterns are listed, together with some relative examples literally translated:

#### - type ASA (ADJ+N+ADJ)

minuta devastazione erbacea	(minute devastation herbaceous)
lunga notte biologica	(long night biological)
grande gorilla nera	(big gorilla black)
perpetua agitazione ansiosa	(perpetual agitation anxious)
esatto equivalente grafico	(exact equivalent graphic)
aguzzo triangolo dorato	(pointed triangle gilt)
opaco colore serale	(opaque colour evening)
solito grigiore quotidiano	(usual greyness daily)

#### - type SAeA (N+ADJ+CONJ+ADJ)

realtà informe e dissennata	(reality shapeless and mad)
attenzione minuziosa e prolungata	(attention meticulous and prolonged)
precisione nomenclatoria e classificatoria	(precision nomenclatural and classificatory)
immensità tranquilla e vuota	(immensity calm and empty)
mondo frenetico e congestionato	(world frenetic and congested)
presenza ingombrante e vistosa	(presence encumbering and striking)
tempo continuo e immutabile	(time continuous and immutable)

#### - type AAS (ADJ+ADJ+N)

irriducibile sorda evidenza (irreducible deaf evidence)  
 bonari rozzi pupazzi (kind rough puppets)  
 monotone spennacchiate livree (monotonous plucked plumages)  
 rosee variegiate airole (rosy variegated flowerbed)  
 stolidi scacazzanti piccioni (stupid craping pigeons)  
 ferma tonda faccia (firm round face)

- type AAA (ADJ+ADJ+ADJ) climax

maldestra, stonata, irresoluta (clumsy, tone-deaf, irresolute)  
 acre, greve, fradicio (acid, heavy, rotten)

- type AeAeA (ADJ+CONJ+ADJ+CONJ+ADJ) climax

grigia e opaca e arcigna (grey and opaque and sullen)  
 fulminea e duttile e prensile (as quick as lightning and ductile and prehensile)

- type AoA (ADJ+DISJ+ADJ) correctio

aria malefica o insidiosa (appearance harmful or insidious)  
 onda contigua o successiva (wave contiguous or subsequent)  
 frasi allusive o sibilline (sentences allusive or sibylline)  
 aggettivo arcaico o vernacolo (adjective archaic or vernacular)

## References

- Calvino I. 1983, Palomar, Einaudi, Torino.  
 Dardano M. 1978, La formazione delle parole nell'Italiano di oggi, Bulzoni, Roma.  
 Mengaldo P.V. 1994, Il Novecento, il Mulino, Bologna.  
 Mengaldo P.V. 1991, Aspetti della lingua di Calvino, in: La tradizione del Novecento, Terza serie. Einaudi, Torino.  
 Picchi E. 1994, Statistical Tools for Corpus Analysis: a Tagger and Lemmatizer for Italian, in: Martin W. et al (Eds), EURALEX94 Proceedings.  
 Roventini A. 1980, Concordanze delle Cosmiche. Analisi statistica del vocabolario e delle variazioni stilistiche, Tesi di Laurea, Università di Pisa, Facoltà di Lettere e Filosofia.  
 Tekavcic P. 1980, Grammatica storica dell'italiano vol.III Lessico, il Mulino, Bologna.

Salva finestra-

Tagger DBTPR7.PLO

```
-----
{Sost }altezza 1
{Sost }ampiezza 1
{Sost }bellezza 1
{Sost }certezza 2
{Sost }compattezza 1
{Sost }completezza 1
{Sost }compostezza 1
{Sost }debolezza 1
{Sost }fanciullezza 1
{Sost }fortezza 1
{Sost }franchezza 1
{Sost }freschezza 1
{Sost }giovinezza 1
{Sost }giustezza 2
{Sost }incertezza 4
{Sost }insicurezza 2
{Sost }leggerezza 2
{Sost }lucentezza 2
{Sost }lunghezza 1
{Sost }naturalezza 1
{Sost }piccolezza 1
{Sost }prontezza 2
{Sost }raffinatezza 2
{Sost }ricchezza 3
{Sost }rilassatezza 1
{Sost }riservatezza 1
{Sost }saggezza 2
{Sost }salvezza 1
{Sost }snellezza 1
{Sost }sostenutezza 1
{Sost }stranezza 1
{Sost }tristezza 1
- DBT -- E:Picchi-----
```

Fig. 1

Salva finestra - Tagger DBTPRI.PLO

{Agg }	acustico	2
{Agg }	antico	1
{Agg }	araldico	1
{Agg }	arcaico	1
{Agg }	aristocratico	1
{Agg }	armonico	2
{Agg }	asimmetrico	1
{Agg }	astigmatico	1
{Agg }	atavico	2
{Agg }	biologico	1
{Agg }	carico	2
{Agg }	cosmico	1
{Agg }	cronologico	1
{Agg }	disarmonico	2
{Agg }	egocentrico	1
{Agg }	fisico	1
{Agg }	frenetico	1
{Agg }	genetico	1
{Agg }	geometrico	1
{Agg }	grafico	1
{Agg }	identico	1
{Agg }	logico	2
{Agg }	meccanico	1
{Agg }	microscopico	1
{Agg }	neoplatonico	1
{Agg }	pratico	1
{Agg }	prospettico	1
{Agg }	ritmico	1
{Agg }	rustico	1
{Agg }	scorbutico	1
{Agg }	simbolico	1
{Agg }	sintetico	1
{Agg }	specifico	1
{Agg }	storico	1
{Agg }	unico	6

- DBT -- E.Picchi -----

Fig. 2

D.B.T. (E.Picchi)

Palomar

abbandonare V/S3IP  
 Forma : abbandona Freq. : 2  
 1) Copito de Nieve» non \*\* abbandona \*\* mai il copertone. Rif.: R077LRNN83X.45.Pag.0083.21  
 2) e si risponde non l' \*\* abbandona \*\* mai: una variazione Rif.: R077LRNN83X.50.Pag.0119.26  
 2

abbattere V/S3IP  
 Forma : abbatte Freq. : 1  
 1) a che non s' \*\* abbatte \*\* sulla riva: e dove Rif.: R077LRNN83X.140.Pag.0019.17  
 1

abitare V/S3IP  
 Forma : abita Freq. : 2  
 1) io egocentrico e megalomane che \*\* abita \*\* in lui. Ma Rif.: R077LRNN83X.25.Pag.0016.1  
 2), il gecko che \*\* abita \*\* sotto le foglie su quel Rif.: R077LRNN83X.12.Pag.0058.15  
 2

abusare V/S3IP  
 Forma : abusa Freq. : 1  
 1) gli abusi di chi \*\* abusa \*\*, egli non ha esitazioni Rif.: R077LRNN83X.116.Pag.0113.27  
 1

accentuare V/S3IP  
 Forma : accentua Freq. : 1  
 1) capta quasi altri dettagli e \*\* accentua \*\* l' astrazione geometrica dell' Rif.:  
 R077LRNN83X.60.Pag.0040.3  
 1

accettare V/S3IP  
 Forma : accetta Freq. : 1  
 1) presunta compagna, che lui \*\* accetta \*\* senza provarla. Rif.: R077LRNN83X.10.Pag.0101.13  
 1

acchiappare V/S3IP  
 Forma : acchiappa Freq. : 1  
 1) La trascura? No, \*\* acchiappa \*\* anche quella. La Rif.: R077LRNN83X.103.Pag.0061.14  
 1  
 .....

-----

Fig. 3

-----

Andromeda W/  
 Forma : andromeda                    Freq. :            1  
 1) Nebulosa del Granchio o l' addensarsi d' un ammasso globulare in \*\* Andromeda \*\* non possono non avere una qualche influenza sul funzionamento del    Rif.:    R077LRNN83X.52.Pag.0119.28  
 1

ArtNouveau W/  
 Forma : artnouveau                   Freq. :            1  
 1) ripiani di questa vetrina allineata una collezione di vasi \_ \*\* ArtNouveau \*\* ó; la sera una lampadina da 75 watt illumina gli    Rif.:    R077LRNN83X.8.Pag.0058.11  
 1

Arturo W/  
 Forma : arturo                            Freq. :            2  
 1) verde. Le stelle intorno sono tutte impallidite, tranne \*\* Arturo \*\* che brilla con aria di sfida un po' pi in alto    Rif.:    R077LRNN83X.21.Pag.0038.24  
 2) quasi ad accucciarsi sulle chiome degli alberi a nordovest; \*\* Arturo \*\* cala a picco sul profilo della collina trascinando tutto l' aquilone    Rif.:    R077LRNN83X.59.Pag.0046.3  
 2

Australia W/  
 Forma : australia                        Freq. :            2  
 1) le nuvole; una si stabilizza in forma e posizione d' \*\* Australia \*\*, e il signor Palomar si convince che pi distinta vede    Rif.:    R077LRNN83X.40.Pag.0039.16  
 2) e il signor Palomar si convince che pi distinta vede quell' \*\* Australia \*\* pi l' obiettivo a fuoco, ma nello stesso tempo    Rif.:    R077LRNN83X.42.Pag.0039.18  
 2  
 .....  
 -----

Fig.4