

Dream-coding project: computer-aided text analysis applied to psychopathology¹

Marco Zanasi (Università di Roma Tor Vergata [marco.zanasi@uniroma2.it]) **Sergio Pizziconi** (Università per Stranieri di Siena [sergio.pizziconi@istruzione.it]) **Daniele Silvi** (Università di Roma Tor Vergata, [silvi@lettere.uniroma2.it])a

1. Preliminaries

A dream is a psychic phenomenon, whose nature is extremely complex and crucial for the understanding of body-mind relationship.

Until Aserinsky and Kleitman (1953) revealed neurophysiological correlates of dreams (REM phase); these had been regarded as an exclusively mental phenomenon, driving researches mainly into the psychological field, thus raising momentum around psychoanalysis and its theories, which formed the foundation of the psychodynamic approach.

Since 1895, psychoanalysis has been dealing with dreams and has used their interpretation to support Freud's Theory and to give evidence of the relevant strength of the unconscious. Neurosciences started coping with dream only later, increasing the understanding of neural processes involved in generating oneiric activity.

The large amount of data gathered in recent years disclosed the existence of specific neural networks involved in the construction of a dream as a whole and in determining some peculiarities, which made it possible to free oneiric contents from outer influences having a mental nature. Some of the characters of a dream are thus organically predetermined, in opposition to the widely spread conception that its genesis and characters are phenomena derived exclusively from a psychological basis.

The network generating oneiric activity is composed of the associative areas and of limbic and paralimbic structures, including significant portions of cognitive areas allowing dramatization and narrative capabilities to present oneiric tales, which usually sketch the dreamer's conscious conceptions and concerns. Dreaming's neural network supplies for the level of brain activation necessary to dreams, and this network is likely to predetermine the type of possible cognitive activation as well (therefore affecting the thematic repertoire of dreams). It is also responsible for the degree of perceptive and emotional intensity, for the sharpness of the oneiric experience. Finally, the neural network can also determine the

peculiar characters of dreams such as lack of selfreflexivity and narrative incongruities. Yet the neural network of oneiric activity cannot account for neither the narrative nature of dreams nor the oneiric contents which are the outcome of the conceptual system of dreams. Many studies and laboratory research agree that the contents of dream reports, both in REM and NREM phases, are mostly represented by a coherent, reasonable simulation of the real world.

The most general conclusion would be that dreams are relatively autonomous and isolated mental phenomena, not being easily affected by induction or modified by before-sleep manipulation. Recognition of neurophysiological structures entitled to the organization of oneiric activity (frontal brain associative cortex, temporal-occipital associative areas) increased our understanding of the phenomenon, but cannot exhaustively explain “import” and “function” of the individual’s oneiric activity. Jung’s hypothesis of the dream as selfrepresentation seems to earn reliability among others: A dream would be a wide open window to the deep psyche, a sensible, precious indicator of alterations in the deepest structures of the psyche and the soma.

The Jungian view makes no difference between latent and manifest meanings. The apparent unreadability of dream is related to the allegorical, allusive, opaque language being used by the unconscious in the oneiric narrative. Thus, a dream is to be considered as a real literary text to which analyses on style, narratology, and structure can be applied. If we consider a text as a woven object, as its etymology suggests, all the more, then, the dream report is to be considered a particular text that turns the oneiric experience into an objective product. A verbal report of the oneiric datum is the only available means to represent the oneiric experience. An oneiric narrative would be impossible if the dream itself were not a text. If we consider a text as a set of woven elements extracted from a specific code and organized in order to communicate something (Danesi 1999), when we recall a dream, either telling it or not, we compose a text, with beginning, central, and closing sections, with formally defined boundaries in time and space; namely, it has a narratological coherent structure. Several distinguished threads of our mind compose the dream and are woven in the loom of the narrative process, which is a part of human cognition. A dream is a text that is similar to a stage setting and reflects the narrative structure of the oneiric experience. The word *narrative* refers to a perceived sequence of events linked together not randomly (Toolan 1988). As suggested by Chatman (1978) in any narrative, there can be distinguished a *story* and a *discourse*². Put simply, the former is the chain of events and tells us “what” the

narrative is about; the latter is the linguistic realization, the expression of the plot, so it is centred upon “how” it is told.

Many studies³ have proved that dream reports are faithful representations of dreams on the basis of at least two types of evidence: Dream contents display strict analogy with waking thoughts; stimuli provoked during sleep are embedded in the oneiric activity and reported. If, following Todorov, a narrative is somewhat like moving from a state of balance to another, separated by a lapse of time in an unbalanced condition, then the dream must be a stable unbalanced state, without a thematic or functional concern neither centred on nor solving any problem. The manifest oneiric image is the dream itself and contains the whole meaning of dream.

Naturally, the transposition of the original oneiric experience into a system of linguistic representation, leads to the intersemiotic translation issues: What is retained and what is transformed? Can we identify in the dream report unmodified invariants of the oneiric experience? Assuming that we can, how could they be linked to the dreamer’s deepest psychic motions if they have been definitively modified by the verbal filtre used in producing the dream report?

We believe that by applying text analysis to the dream as a “graphical” expression of an unconscious language, it will be possible to extract important pieces of information from material apparently as chaotic and unstructured as dreams.

2. Dream Coding and Project Phases

2.1 Dream coding

The goal of “Dream Coding” is to collect, transcribe, catalogue, file, study, and comment on dreams reported by selected patients at Tor Vergata University Hospital in Rome. In the group of observed patients, different psychopathological statuses are represented.

2.2 Project phases

The work is complex by its nature and is therefore structured in several phases.

2.2.1 Data-collection

Patients who have already a precise diagnosis are called “type A”. The data collected in the table shown in Fig. 1 are socio-demographics, clinical and linguistic data.

Fig. 1 Patient's data table

SCHEDA ANAGRAFICA	
ID CASO	_____
COGNOME	_____
NOME	_____
SESSO	DATA DI NASCITA
LUOGO DI NASCITA	_____
TITOLO DI STUDIO	_____
OCCUPAZIONE	_____
DIAGNOSI	_____

STATUS	_____

DATA RICOVERO	DIALETTOFONO

An “ID case” number is used to guarantee the anonymity of the patient. This identifier is necessary in order to use the data freely without constraints of any bias. The following information is also of paramount importance in determining linguistic framework: city of birth (or region), whether the patient speaks dialect, the highest education accomplished, and profession. These data are collected for patients of Type A (where the diagnosis is known). The goal in the first part is to correlate linguistic features to specific pathologies. In the second part, the goal is to transcribe, catalogue and file dream reports using the table in Fig. 2.

Fig. 2 Dream-report table

SCHEDA SOGNO		
ID CASO	_____	
ETÀ	_____	GIORNO SOGNO _____ NUM. SOGNO _____
DIAGNOSI	_____	

STATUS	_____	

TESTO DEL SOGNO		
.....		
.....		
.....		
.....		
.....		

An ID case is used in this table also. The ID is necessary to have a link between the two archives. The general data recorded for each patient will have the ID number repeated in the tables of his/her dream reports. It is important to emphasize that the two archives need to be separated in order to have the certainty that some of the analyses conducted are not influenced by demographic and social aspects of the patients.

Dream collection can be divided into three different phases:

- a. audio-recording,
- b. transcription,
- c. filing.

During the first phase, dreams are recorded and then transcribed as accurately as possible by an operator using the table in Fig. 2. The filing is both on paper and on tapes.

2.2.2 Dream analysis

After collecting and filing the material to be analysed (dream reports of Type A patients), the study continues with the coding of the dreams.

The activities of encoding and marking have been designed to meet the requirements below:

- a. Source texts are spoken; therefore, methodologies and techniques used in building corpora of spoken Italian will work as a reference framework⁴.
- b. Modelling needs to be stressed because slight differences can lead to a totally different conclusion.
- c. Revising and amending coding protocol should be always possible according to the results obtained.

Once the model has been chosen, texts can be analysed using a software program that processes both statistical and lexical research. One of these programs is called TAPoR, which allows one to query both plain text and XML tags.

TAPoR stands for *Text Analysis Portal for Research*, and, as the name suggests, it performs text analysis. TAPoR is the result of the collaboration of six Canadian universities: McMaster University, University of Victoria, University of Alberta, University of Toronto, University of Montreal, and University of New Brunswick. TAPoR uses collective data gathered by the collaboration of the previously mentioned six universities. The project has its own roots in the work of Tim-Berners Lee, director of the *World Wide Web Consortium* and inventor of the Web. He believes in the creation of a *Semantic Web*⁵—simply put, a coding procedure aimed at producing intelligent, intelligible texts, which can be analyzed automatically by software like TAPoR. To some extent the semantic web would generate an operational definition of texts—restricting them to those documents that can be digitally represented.

Apart from being a new reference software for computer-aided calculation in text analysis and involved sectors, TAPoR provides the opportunity for the community of humanists to develop models, instruments, and methods which will mark the path for the next decades, taking part actively in the project. TAPoR analyses three kinds of texts: plain text (.txt), HTML, XML.

At this stage of the project, the analysis will correlate linguistic features and psychopathologies on a statistical, aggregated basis. This is the first step toward the definition of protocols of analysis of single patient's linguistic realizations.

2.2.3 Testing the protocol

The protocol created in the previous phase will eventually be verified and corrected. To achieve this, it will be applied with patients with unknown diagnoses (identified as Type B patients). Starting from the linguistic hypotheses, the individual language use will be regarded as an indicator of which is the most likely psychopathology from which the patient suffers. To

reach complete verification, the analysis protocol will be applied adding step by step more complex variables.

2.2.4 Results comparison

In this last phase, Type A and Type B patients are compared. At this point a grid that shows linguistic features linked to single pathologies is created. Obviously, the study starts with the linguistic peculiarities that have a higher correlation.

At this point, a “linguistic diagnosis” to associate with a “clinical diagnosis” is proposed.

2.3 The scale of silence

Patients who tell dreams usually pause from time to time. We are not referring to the kind of pausing in spoken realization which is paralleled by the different signs of punctuation in written language. Most likely because they are not telling their dreams spontaneously, but after being prompted to, they often stop—probably recalling the images of their oneiric activities. Obviously it would be hard to make a clear, sharp difference between these two kinds of pausing. Therefore we have decided to create symbols different from the usual annotation of short, medium, and long pauses. The new set of symbols is shown in Fig. 3. All usual prosodic pauses will be marked as a *short pause* (#), while a longer silence—which is the kind that is more specific to dream reporting—will be marked by (##) and (###) according to duration.

Fig. 3 Scale of silence

Short Pause 0-2 sec (#)	Medium Pause 2-4 sec (##)	Long Pause 4+ (###)
----------------------------	------------------------------	------------------------

3. Linguistic Approach

3.1 Theoretical and methodological issues

This research project represents more than a challenge to the linguist for the theoretical and methodological implications linked to the definition itself of the object⁶ of study. On one hand, it is clearly different from those language deviations (e.g. aphasia, acquired dyslexia) when they are determined by injuries or any other physical or biochemical change in brain physiology. On the other hand, the object of study must be independent from variables (e.g. time, space, gender, age, social groupings, and professional and work activity) that characterize the variations of language typically studied in sociolinguistics. The latter, even in

literary-style studies, are always somehow focused on the intersubjectivity that affects speakers' (writers') personal histories, social backgrounds, processes of socialization, educations, and peers. On the contrary, our object of study is variants of language that must not be affected by these variables and must correlate only to the speaker's specific clinical status. The level to be investigated would, hence, fall between these two ends⁷.

Nevertheless, the group of speakers we observed in the preliminary stage of the research still speak a language which superficially appears as a normal (statistically meant⁸) spoken Italian. As we will see below going through few examples of dream reports gathered and transcribed by the Psychiatric Unit of Tor Vergata, they display neither unparalleled syntactic structures nor fancy lexical choices. But if it is true that, for instance, schizophrenic patients miss the ability to interpret and produce a metaphoric use of language, there might be linguistic features making the difference.

What are the linguistic features which can account for the variation between dream reports of people who suffer from psychopathologies and those of people who do not? And what are the features characterizing different psychopathologies? How should these features be assessed? Finally and consequently, how should they be marked to prepare corpus for analysis?

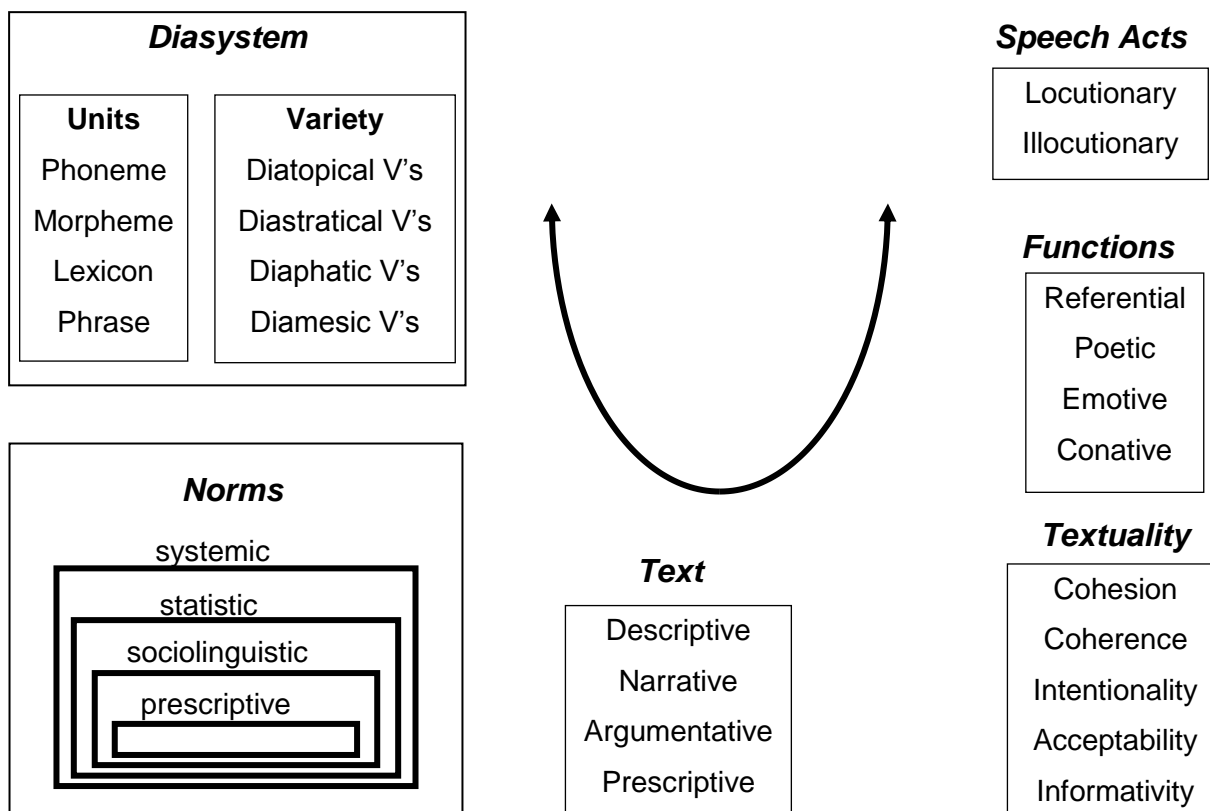


Fig. 4 The diatextual rule (Catricalà, 1995 and 2004)

The set of variables that could be chosen is as wide as the multilayered, multidimensional process that leads a speaker to realize any utterance. The scheme of the diatextual rule (Fig. 4), translated and adapted from Catricalà (1995 and 2004), helps us to understand the interrelated choices to go through when in a specific situation a text is to be produced. For instance, we may want to compose a message slightly marked on the diatopical⁹ axis since we want to select a sociolinguistic norm widely used (highly ranked in the statistical norm, but out of the prescriptive one) to give our narrative a deeper emotive and conative nuance. This would make us move in the diasystem¹⁰—choosing the most appropriate variants at the phonological, morphological, lexical, and syntactical level, stressing the perlocutionary aim of speech, while playing on the acceptability of the text¹¹.

The diatextual rule does not only display the width of the scope of linguistic variables that could be taken into consideration¹², but also it shows as well that the links between variables are as important as the variables themselves. For some variables the quantitative results are expected to be ordered

in data arrays like the one showed in Table 1 and comparable to those of the *Varietätengrammatik* of Klein and Dittmar¹³. The first column gives a block of rewrite rules all having the same entry

Table 1 Sample data array					
	Psychopathologies				Normal
	1	2	3	4	Italian
R → x	p _{1x}	p _{2x}	p _{3x}	p _{4x}	p _{sx}
R → y	p _{1y}	p _{2y}	p _{3y}	p _{4y}	p _{sy}
R → z	p _{1z}	p _{2z}	p _{3z}	p _{4z}	p _{sz}
R → w	p _{1w}	p _{2w}	p _{3w}	p _{4w}	p _{sw}

and different exits recorded in the corpus. Each rule is provided with relative frequencies calculated in the subcorpora of the dreams reported by patients grouped according to their psychopathologies.

However, some variables will be registered in a dynamic way according to the three modalities below.

1. Marking a switch from one value to other(s). Concerning verb tenses, when switches occur from present to past and vice versa—see sample text (2) below. At the syntactical level, when switches occur from passive to active diathesis. As for textual structures, when switches occur from direct to indirect speech and to the so-called free indirect speech. Still at the textual level of analysis, anaphora can occur in each sentence either by repetitions, substitutions, pro-forms, or ellipses. Switches from one device to other will be marked. The use of ellipses, especially in the subject position, is particularly significant to the research objective, since Italian is a PRO-drop language.
2. Marking co-occurrences of specific values in different variables. For instance, it might be meaningful to register the co-occurrence of passive diathesis and repetition of subject personal pronoun *io* (*I* in English), which would be usually omitted, to stress a particular representation of agentivity in the sentence. At different levels of analysis, we could mark the co-occurrence of intonation variation and lexical choices.
3. Marking co-occurrence of specific values in a linguistic (*strictu sensu*) variable with occurrence of specific values in other textual and paralinguistic variables. The latter is, for instance, emphasis; the former is either the content categories used to analyse dream reports and listed by Domhoff and Schneider (1998)¹⁴, or *figurative units*¹⁵, or so-called *stanzas* already used in discourse analysis of narrative (Gee 1998).

3.2 Few marking issues

In this section the marking of some linguistic features to be annotated in corpus preparation will be described. They are not the only items that will be kept into consideration and have been chosen by the research team to account (a) for distinct levels of analysis and (b) for some of the problems raised by the transcription and annotation of this peculiar corpus of spoken Italian¹⁶.

The first example of dream report, showing the use of the symbols in the scale of silence, displays how prosody affects syntactic assessment. In fact, in (1) two *che* (*that*), with a double underline, are to be evaluated as fillers and not as conjunctions. The presence of a pause before them supports this evaluation, turning what could be interpreted as a hypotactical structure into a juxtaposition. In the first stage of the research, features characterizing spoken language, such as fillers and false starts (single underline), will be marked with the clear intent to help performing other evaluations. Even though prosodic peculiarities, such as intonation variation and emphatic pronunciation, will be signalled to establish co-occurrences with other linguistic features.

Sample (1)

**Ho sognato che # andavo al mare # c'era una festa ## o una cosa del
[I] dreamed that [I] was going to the sea there was a party or one thing of the
genere ### il sogno di questa notte è che andavo a questa festa # che incontravo
kind the dream of this night is that [I] went to this party that [I] met
tutti i miei parenti solo da parte di mio padre ## però alcuni non li riconoscevo #
all my relatives only from the side of my father but a few [I] didn't know
altri # ehm # mi ricordo che c'era questa casa # c'era questa casa
others ahem [I]remember that there was this house there was this house
dove c'era questa festa # che era una casa popolare
where there was this party that was a house poor
quindi però era molto grande
hence but [it]was very large**

Sample (2) shows verb-tense switching from indicative imperfect to indicative present. This variable has proved to be correlated to the respective ages of those reporting their dreams. Zanasi, De Persis, Caporali, and Siracusano (2005) revealed that the old either use present tense or past tense but seldom switch, while the young never stick to one tense and always switch.

Sample (2) [verb tenses: present; imperfect]

**Camminavo per strada e incontro mia nipote
[I] was walking in [the] street and meet my niece
che non vedo da tanto tempo**

whom [I] don't see from long time

ed era con un'amica e aveva i capelli tutti lunghi mossi castani

and [she] was with a friend[FEM.] and had hair very long wavy brown

era giù e io le chiedo perché [...]

[she] was down and I her ask why [...]

Sample (3) shows three anaphoric threads interwoven in the dream of a woman in her fifties. The selection of pro-form, repetition, substitution, and ellipsis might prove significant to the object of study, not only for what was stated in points 1 and 2 in Section 3.1 above. First, we have observed that repetitions by some patients seem to be the almost exclusive option. Second, signalling the endophoric structure will help to evaluate cohesion and coherence in the text.

Sample (3) [referent: dreamer, patient's niece, patient's daughter]

[io] Camminavo per strada e [io] incontro mia nipote che [io] non vedo da tanto

[I] was walking in [the] street and [I] meet my niece whom [I] don't see from long

tempo ed [lei] era con un'amica e [lei] aveva i capelli tutti lunghi mossi castani

time and [she] was with a friend[FEM.] and [she] had hair very long wavy brown

[lei] era giù e io le chiedo perché [lei] piangeva e lei mi risponde che [lei]

[she] was down and I her ask why [she] wept and [she] to me replies that [she]

piangeva perché Erica non andava mai a trovarla e allora io cercavo di giustificare

cried because Erica never went to call her and so I tried to justify

Erica perché lei va a scuola poi il pomeriggio [lei] lavora quindi [io] cercavo di trovare

Erica because she goes to school then in the afternoon [she] works thus [I] tried to find

giustificazioni al fatto che lei e la cugina non si vedevano e

justifications to the fact that she and the cousin did not see and

a me urtava questa cosa che lei piangesse.

to me this thing hit that she cried

Discrepancies in cohesion and coherence might, of course, be justified by the fragmented nature of contents in both the experience and the report of dreams. Yet, as shown in the other sample texts, they do not seem to be so frequent and everywhere as they are in sample (4).

Sample (4)

Io mi ricordo praticamente come se io visitassi varie stanze

I remember (say like) as if I visited several rooms

dove vedevo dei cadaveri dove il mio corpo mi sembrava fossi io morta

where [I] saw some corpses where my body seemed to me it was I dead

insomma però io ero viva e vedevo questi ## questi miei corpi morti.

well but I was alive and saw these these my bodies dead.

The last two examples show the use of metaphoric discourse with two different meanings. In sample (5) a metaphor is used in an odd way. The analogy between the swarm of bees and the attack of the dolls is not based on the central, prototypically speaking, features

of the number of elements involved. This patient refers probably to the kind of alternating movement of bees flying around a flower, which is a feature not as central as the former.

Sample (5)

Ho avuto un sogno e c'era una coppia di bamboline

I had a dream and there was a pair of little dolls

come uno sciame di api che si avvicinava vicino vicino a me e

like a swarm of bees that came over close close to me and

il diavolo ha detto io prendo te [...]

the devil said I take you [...]

Instead, in sample (6) we have marked four examples of the type of metaphors that Lakoff and Johnson (1980) regard as central in the connection between mind, body, and language.

Sample (6)

Cammino per strada e incontro mia nipote che non vedo da tanto

[I] was walking in [the] street and [I] meet my niece whom [I] don't see from long

tempo [...] era giù e io le chiedo perché piangeva [...]

time [...] [she] was down and I her ask why [she] wept [...]

io cercavo di giustificare Erica [...] cercavo di trovare giustificazioni al fatto

I tried to justify Erica [...] [I] tried to find justifications to the fact

che lei e la cugina non si vedevano e

that she and the cousin did not see and

a me urtava questa cosa che lei piangesse.

to me this thing hit that she cried

In the first metaphoric element, single underline, the emotional state of the dreamer's niece is described referring to the scheme that HAPPY IS UP, SAD IS DOWN. "Trovare giustificazioni" refers to the scheme that CAUSES AND EXPLANATIONS ARE HIDDEN OBJECTS. In the third example, dotted underline, the use of the verb *vedere* (to see), meaning 'to meet', pertains to the scheme that SEEING IS TOUCHING. The fourth and last example shows the use of the verb *urtare* (to hit) meaning 'to annoy, irritate, disturb', brings up to the scheme that EMOTIONAL EFFECT IS PHYSICAL CONTACT.

4. XML Coding: a TEI Dream?

As affirmed before, the literature on the corpus design of spoken language will be retained as a general framework in order to give the material gathered the maximum intelligibility in the dissemination of the results phase. A few issues in the transcription and processing of spoken data will be related in the following to account for the peculiarities of the material of study. It

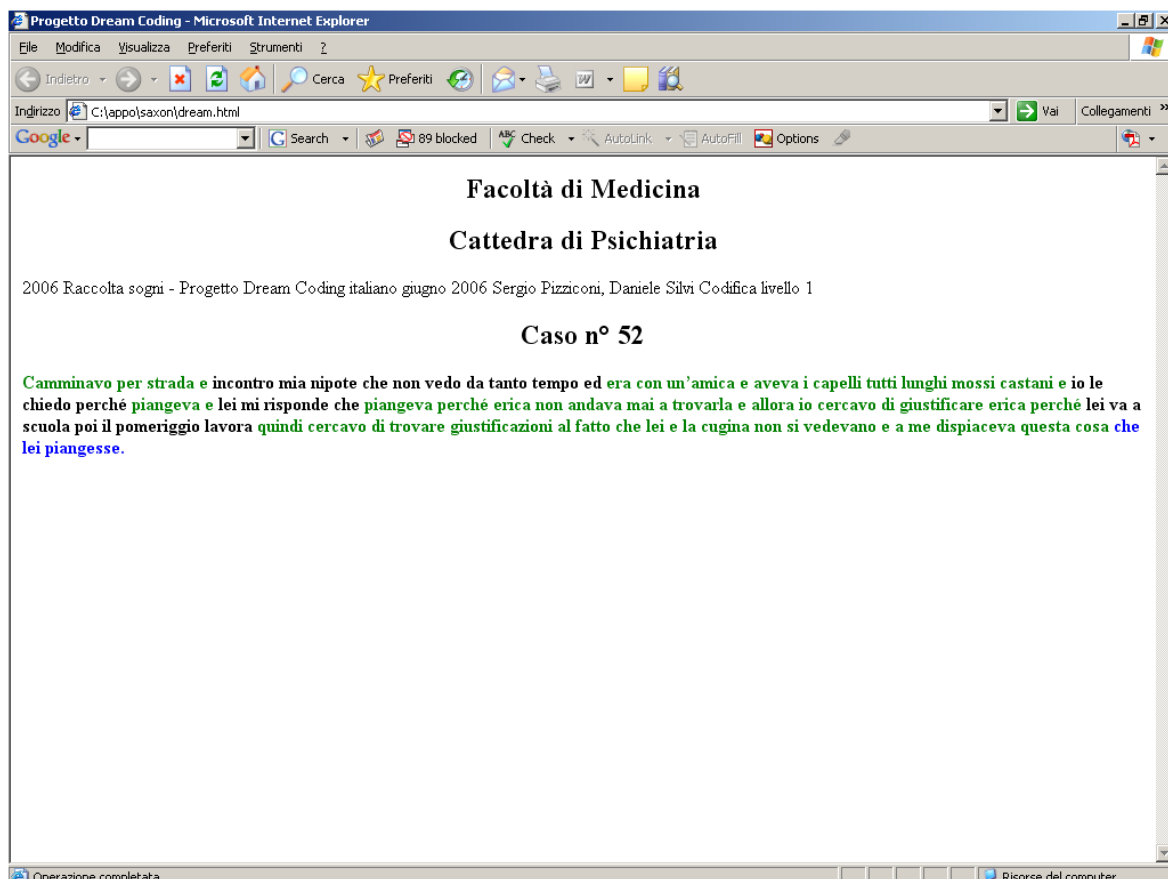
should be clear that in this early stage of the research, any proposed solution is to be considered as a provisional one, a starting-step toward the verification of linguistic and psychological hypotheses. In the light of the first findings, the tags and rules of coding will be surely emended and better adapted to dream reports.

The first coding will be based on TEI P4 release, using the tagset called *Interpretation and Analysis*. Some of the tags to be used are: <interpGrp>, <interp>, <s>; among attributes: ana, type, and id. XML coding devices that have proved to be efficient such as to mark some of the linguistic items discussed previously are:

- Intonation features: tag
- Verb morphosyntax: tag
- Pro-forms, repetitions, substitutions, and ellipses: tag with label of cross-reference
- Metaphors: tag
- Pauses: entities

To display texts an XSLT transcription has been chosen because it provides us with query structures inside tag and attributes. Furthermore, its syntax is equivalent to XML's; this similarity makes the work easier. An XSL style sheet applying to all coded phenomena has been implemented. A short example of the view is presented in Figure 5.

Fig. 5 Sample style sheet



In the example (Fig. 5) the moods and tenses of verbs have been highlighted with different colours: Black is for the indicative simple present; green is for the indicative past (in this case, imperfect); and blue for the subjunctive mood. These colour codings also clearly show the switches. A few lines of the XML source text are shown below:

```
<seg type="mc" rend="imp">Camminavo per strada</seg>
e <seg type="verbt imperf" rend="para">
incontro mia nipote </seg>
<seg type="verbt pres" rend="hypol">che non vedo</seg>
da tanto tempo ed<seg type="verbt imperf" rend="hypol">era
con un'amica e aveva i capelli tutti lunghi mossi castani
e</seg>
```

5. The Analysis with TAPoR

As briefly stated above TAPoR is designed to analyse texts not only in a descriptive way, but also it allows the researcher to profit extensively from a coding which is also semantic. The portal has been referred to as a gateway to several tools for subtle, deep text analyses. It supplies the researcher with sample texts to be used in experimental stages. Texts cover a wide scope of topics such as history, law, literature, etc.

Other tools specifically designed to process spoken texts in preparing our corpus are listed below:

- Voice Walker™ (free)
- Sound Writer™ (free)
- TranScriber™ (gnu)
- SoundScriber™ (free)
- TAPoR™ (on line, free)

References

Albano Leoni, F. (2003). Tre progetti per l'italiano parlato. In Maraschio, N. and Poggi Salani, T. (eds) (2003). *Italia linguistica anno mille. Italia linguistica anno duemila*. Atti

- del XXXIV Congresso Internazionale di Studi della Società di Linguistica Italiana. Roma: Bulzoni, pp. 675-84.
- Albano Leoni, F., Cutugno, F., Pettorino, M. and Savy, R. (eds) (2004).** *Il parlato italiano. Atti del Convegno Nazionale*. Napoli: D'Auria Editore.
- Aserinsky, E. and Kleitman, N. (1953).** Regularly occurring periods of eye motility and concomitant phenomena, during sleep. *Science*, 1 118: 273-4.
- de Beaugrande, R. and Dressler, W. (1981).** *Introduction to Text Linguistics*. London: Longman.
- Catricalà, M. (1995).** *L'italiano tra grammaticalità e testualizzazione*. Firenze: Accademia della Crusca.
- Catricalà, M. (2004).** *Forme, Parole e Norme. Lineamenti sociolinguistici dell'italiano contemporaneo*. Milano: FrancoAngeli.
- Coseriu, E. (1952)** Sistema, norma y habla. In Coseriu, E. (1952) *Teoría del lenguaje*. Madrid: Gredos
- Courtés, J. (1986).** *Le Conte populaire: poétique et mythologie*. Paris: Presses Universitaires de France.
- Cresti, E. (2000).** *Corpus di italiano parlato*. Firenze: Accademia della Crusca.
- Cresti, E., Moneglia, M. et alii (2002).** The C-ORAL-ROM Project. New methods for spoken language archives in a multilingual romance corpus. In Rodriguez, M.C. and Suarez Araujo, C. (eds) (2002), *Proceedings of III International Conference on Language Resources and Evaluation*. Paris: ELRA, Vol 1, pp. 2-10.
- Danesi, M. (1999).** *Of cigarettes, high heels and other interesting things*. New York: St. Martin's Press.
- De Mauro, T., Mancini, F., Vedovelli M. and Voghera M. (1993).** *Lessico di frequenza dell'italiano parlato*. Milano: Etass Libri.
- Dittmar, N. (1989).** *Variatio delectat. Le basi della sociolinguistica*. Galatina: Congedo.
- Domhoff, G.W. and Schneider, A. (1998).** New rationales and methods for quantitative dream research outside the laboratory. *Sleep*, 21, 398-404.
- Firenzuoli, V. (2000).** Nuovi dati statistici sull'italiano parlato. *Romanisches Forschungen*, 13: 213-225
- Gee, J.P. (1998).** Two styles of narrative construction and their linguistic and educational implications. In Cheshire, J. and Trudgill, P. (eds.) (1998). *The Sociolinguistics Reader. Gender and Discourse. Vol. 2*. London – New York: Arnold, pp. 295-314.

- Hjelmslev, L.** (1943). *Omkring sprogteoriens grundlæggelse*. Copenhagen: Akademisk forlag.
Italian translation: *Lingua e parole*. In Hjelmslev, L. (1981) *Saggi di linguistica generale*.
Translated and commented by Massimo Prampolini. Parma: Pratiche Editrice, pp. 91-114.
- Klein, W.** (1988). Varietätengrammatik. In Ammon-Dittmar-Mattheier (eds),
Sociolinguistics/Soziolinguistik, Vol. II. Berlin – New York: de Gruyter, pp. 997-1006.
- Klein, W. and Dittmar, N.** (1979). *Developing Grammars. The Acquisition of German by Foreign Workers*. New York – Heidelberg: Springer.
- Kramer, M.** (1993). Dream translation: an approach to understanding dreams. In Delaney, G. (ed.) (1993). *New directions in dream interpretation*. Albany, NY: SUNY Press, pp. 155-94.
- Lakoff, G. and Johnson, M.** (1980), *Metaphors We Live By*. Chicago – London: The University of Chicago Press.
- Mioni, A.M.** (1983). Italiano tendenziale: osservazioni su alcuni aspetti della standardizzazione. In *Scritti linguistici in onore di Giovan Battista Pellegrini*. Pacini: Pisa, pp. 495-517.
- Moneglia, M.** (1999). Le corpus LABLITA. In Bilger, M. (ed.) (1999). *Corpus. Méthodologie et applications linguistique*. Paris; Champion, pp. 49-57.
- Moro, A.** (2006), *I confini di Babele. Il cervello e il mistero delle lingue impossibili*. Milano: Longanesi.
- de Saussure, F.** (1922). *Cours de linguistique générale*. Paris: Edition Payot.
- Simone, R.** (2006a). Constructions: types, niveaux, force pragmatique. In Prévost, S. (ed.). *Mélanges pour Christiane Marchello-Nizia*. Paris: Ecole Normale Supérieure.
- Simone, R.** (2006b). Nominales sintagmáticos y no-sintagmáticos. In De Miguel, E. *et al.* (eds), *Estructuras léxicas y estructura del léxico*. Berlin: Peter Lang, pp. 225-246.
- Sornicola, R.** (2005). Italiano parlato, dialetto parlato, parlato. In Burr, E. (ed.), *Tradizione & Innovazione I: Il parlato: teoria - corpora - linguistica dei corpora*, Atti del VI Convegno SILFI, (Gerhard-Mercator Universität Duisburg 28.06.-02.07.2000), Firenze: Cesati, pp. 25-39.
- Toolan, M.J.** (1988). *Narrative: a critical linguistic introduction*. London – New York: Routledge.
- Todorov, T.** (1969). *Grammaire du Décameron*. The Hague: Mouton.
- Voghera, M.** (1992). *Sintassi e intonazione nell'italiano parlato*. Bologna: Il Mulino.
- Voghera, M.** (2001). Teorie linguistiche e dati di parlato. In Albano Leoni, F., Stenta Krosbakken, E., Sornicola R. and Stromboli, C. (2001). *Dati empirici e teorie linguistiche*. Roma: Bulzoni, pp.75-95.

Weinreich, U. (1953). *Languages in Contact*. The Hague: Mouton.

Zanasi, M., De Persis, S., Caporali, M. and Siracusano, A. (2005). Dreams and Age. *Perceptual And Motor Skills*, 100: 925-38.

¹ The contents of this paper have been discussed and organized together by the co-authors. However, section 1 is due to Marco Zanasi, section 3 to Sergio Pizziconi and sections 2, 4, and 5 to Daniele Silvi.

² But the distinction is at least as old as the term *narratology*, coined by Tzvetan Todorov (1969).

³ See Kramer (1993) for other references.

⁴ Karl-Franzens-Universität Graz (Austria) published the BADIP - *Banca dati dell'italiano parlato* (<http://languageserver.uni-graz.at/badip/badip/home.php>), where 19 corpora of spoken Italian are listed at the time of this paper.

⁵ <http://www.w3.org/DesignIssues/Semantic.html>

⁶ Here *object* is to be intended as in the opposition *objet vs matière* that has been outlined in chapters 2 and 3 of the introduction to the *Cours de linguistique générale* (Saussure 1922). Consequently, the transcribed texts of dream reports will be the *matière* of the study, whereas the language variation to be investigated is the *objet*.

⁷ Using a computer-technology metaphor to clarify the meaning of this distinction, we would say that language variations caused by brain injuries affect the hardware and variations across space, social strata, gender, scientific or technical specialization, and so on affect the software. Whereas, the variation to be investigated in our research would affect something in between the first two elements, which we could identify as the operative system. The biochemical alterations linked to some psychopathologies can effectively be taken into account in the metaphor considering the interaction that an operative system has both with the hardware and the software of a computer framework. If the main hypothesis of this research were verified, a wider discussion on language production could be stimulated considering the theoretical import in postulating the presence of this middle stage of language processing, especially under the light of the most recent work by Italian scholar Moro (2006), who gives us evidence on the correlation of language production and specific activation areas of the brain.

⁸ The adjective *normal* will be used in this section exclusively with this meaning.

⁹ The *diamesic* axis, introduced by the Italian scholar Mioni (1983) to point out the differences between written and spoken languages, has been added to the well known set of terms used by Coseriu (1973) to concisely refer to the main types of language variations. Diamesic variation has a wider importance in Italian because of its particular history in which it has been for centuries only a written language.

¹⁰ The diasystem was defined by Uriel Weinreich (1953) as a system within which the speaker has multiple options for every point. Language contacts analysed by the author are not the only cases in which the concept of diasystem earns a significant role. If not extensible to any repertoire, the diasystem can be effectively linked to the status of *lingua cum dialectis* which characterizes Italian language.

¹¹ *Acceptability* is one of the text requirements of textuality listed in the scheme and defined by Beaugrande and Dressler (1981).

¹² It is useful to remark that in the scheme in Fig. 1 I have omitted paralinguistic and extralinguistic features such as prosody, prosemics, and kinesics, which can deeply affect speech. It is also important to notice that the distinction between system and discourse, which turns out to be crucial in pragmatic approaches (Simone 2006a and 2006b), is taken into account in the concentric representation of the normative system. Different relationships between systemic norm and use which determine causative schemes (i.e. what is a function of what) as different as Coseriu's (1952) and Hjelmslev's (1943) have not been marked in the diatextual rule (see Catricalà 1995 and 2004).

¹³ Klein (1988), Dittmar (1989), and Klein and Dittmar (1979). In the research carried out by both the authors the varieties to be described were the interlanguages spoken by four different groups of immigrants in Heidelberg, Germany. In one of the data results the block of rewrite rules in the first column presented the entry NP and the following rewrites: NP→(Mod) N; NP→Det (Mod) N; NP→Quant (Mod) N; NP→Num (Mod) N; NP→(other structures). The term of comparison was, of course, the frequencies of the same NP in a corpus of spoken German recorded among L1 speakers in Heidelberg.

¹⁴ A few examples will clarify the type of categorization. About the characters in the reported dream, the authors list: animal, male/female, familiarity, friends, and family; about settings: indoor, familiar setting.

¹⁵ In figurative arts studies, a *figurative unit* is the set of iconic devices used to give shape and colour to the so-called *theme*. Courtès (1986) proved that it is an effective analytical tool in studying popular tales. In a private conversation Maria Catricalà suggested that iconic configuration of figurative units might also be significant in distinguishing psychopathologies. This is one of the qualitative aspects which will be investigated.

¹⁶ Researches in spoken Italian have gained wider and wider interest in several disciplines from sociolinguistics to computational linguistics. It would be impossible to mention all the scholars who have worked on the topic.

References in the following works will give an idea of the field. First, the so-called LIP by De Mauro, Mancini, Vedovelli, and Voghera (1993) is still the most complete work on frequencies of spoken Italian. Information on the wide corpora of spoken Italian is in Cresti (2000), Moneglia (1999), Cresti, Moneglia, et al. (2002) and see note 2 in this paper for references to a list of corpora published on the web. Albano Leoni, Cutugno, Pettorino, and Savy (2004) is the proceedings of the most recent conference wholly dedicated to spoken Italian and a large amount of studies among which only few can be mentioned such as: Voghera (1992 and 2001), Firenzuoli (2000), Albano Leoni (2003), Sornicola (2005).