

NLP Resources for the Analysis of Patient/Therapist Interviews

Horacio Saggion*, Elena Stein-Sparvieri**, David Maldavsky**

*Department of Computer Science
University of Sheffield
211 Portobello Street - Sheffield, England, UK
H.Saggion@dcs.shef.ac.uk

**Instituto de Altos Estudios en Psicología y Ciencias Sociales
Universidad de Ciencias Sociales y Empresariales
Paraguay 1401, PB, Bs. As. Argentina
estein@soluitions.com.ar;dmaldavsky@elsitio.net

Abstract

1. Introduction

In recent years the field of Natural Language Processing (NLP) has seen a renewed interest on the analysis of non-factual, emotional discourse characterized by the presence of affective language and sentiments, and charged with subjectivity. One area which has not been properly investigated, however, is that of natural language processing in the field of psychology and more specifically to the analysis of interactions between patients and therapists. We are developing a set of NLP tools and resources for the analysis of interviews framed on a psychoanalytic theory and the work presented here is the first to investigate the application of NLP techniques for the automatic analysis of spoken transcriptions in Spanish (Argentinian variety) of psychoanalysis sessions between therapists and patients.

The automatic analysis of the sessions which is used as a tool for psychoanalyst to assess and interpret are framed on Freudian theory developed by Libermann and extended by Maldavsky. The automatic tools to be presented here aim at recognizing a subset of Freudian drives manifested in both patient's and therapist's discourse. We are also investigating the applicability of the techniques to other discourses charged with subjectivity.

Abbreviation	Drive Name
IL	Intra-somatic libido
O1	Primary oral
O2	Secondary oral sadistic
A1	Primary anal sadistic
A2	Secondary anal sadistic
UPH	Urethral phallic
GPH	Genital phallic

Table 1: Drives in Libermann and Maldavsky theory

The theory identifies 7 drives which are introduced in Table 1 some interpretation may identify these drives with emotional or affective states such as: strong emotions associated with IL; ecstasy or trance with O1; sadness with O2; anger with A1; etc.

Drive	Lexicalization
A2	verbs: must, to know, to study, to investigate, to possess, to dominate; nouns: vice, doubt, uncertainty, idea, morals, obligation, oath, tradition; adjectives: good, bad, clean, dirty, guilty; adverbs: but, although, however.
UPH	verbs: to be able, to dare, to be accustomed, to cut, to interrupt, to avoid, to hide; nouns: friend, image, scar, precipice, wound; adjectives: coward, scared, tiny, dangerous; adverbs: almost, a bit.
GPH	verbs: to promise, to give, to offer, to receive, to fascinate, to delight, to shine, to seduce; nouns: beauty, ugliness, amazement, ornament; adjectives: wavy, pretty, deformed, huge; adverbs: more, even, besides, mainly, marvelously.

Table 2: Sample of drives A2, UPH, and GPH and associated lexicalization

The theory also associates lexicalizations to each of the drives, they have been carefully selected following a variety of methods:....We show some lexicalizations in Table 2.

2. Language Resources

Text analysis is based the use of a dictionary which has been implemented as a language resource in GATE. It is based on lists of root forms which have been created for each of the drives. The lists are organized according to their parts of speech. An instance of the dictionary is created from the set of 24 lists and kept on-line for processing (human annotation or automatic analysis).

An annotation tool has been implemented using the functionalities provided by the GATE infrastructure. The tool allows a researcher annotate words to be included in the dictionary.

3. Text Analysis

We have implemented a series of programs to process interviews and other types of textual data in Spanish. We are using the GATE system (Cunningham et al., 2002) infrastructure only, most developments are new and they are packaged in a plug-in which can be accessed through the GATE system or standalone. We have developed various programs to automatically annotate the interviews: The following tools have been developed:

- A parts of speech program from the TreeTagger package (See <http://www.clarin.eu/tools/treetagger>) is used to process the document;
- An alignment program has been developed to associate the output of the tagger to the actual text of the interview, therefore creating word annotations;
- A sentence identification program is used to identify sentence boundaries and types of sentences;
- A segmentation program is used to identify patient and therapist interventions;
- A named entity recognizer for Spanish is being developed using Support Vector Machines and training data from the CoNLL evaluation programme;
- An processing resource uses the dictionary and interprets each word according to the drives in the dictionary;
- A processing resource has been implemented to generate an interpretation of the different languages at different segments (therapist or patient or any other segment of interest);
- Statistics are computed for each of the segments.

4. Evaluation

Evaluation of the tools investigated here represent a challenging research question, specially when extrinsic evaluation is considered. Where the statistical distribution of types of languages is concerned, patients discourses can be automatically analyzed by the tools and this result compared with the interpretation given by a therapist. Agreement can be measured...

5. Outline of the Paper

6. Appendix

T: ¿con que te cortaste?	T: What did you cut yourself with?
L: con un vidrio que encontré en el patio	L: With glass I found in the patio.
T: ¿donde lo tenías?	T: Where did you have it?
L: en el locker, en la puertita del locker, y después lo puse en la jabonera cuando baje a bañarme	L: In the locker, in the locker's small door, and then I put it in the soap box when I went down to have a bath.
T: o sea, ya tenías un vidrio escondido	T: That is to say, you already had a hidden piece of glass.
L: sí, ayer lo encontré	L: Yes, yesterday I found it.
T: ¿ayer a la tarde?	T: Yesterday afternoon?
L: sí, sí, de ayer a la tarde	L: Yes, yes, from yesterday afternoon.
T: ¿lo buscaste?	T: Did you look for it?
L: sí, sí lo busqué	L: Yes, yes I did
T: buscando encontraste.	T: Looking you found.
L: ¿eh?	L: Eh?
T: buscando encontrás	T: Looking you find
L: sí	L: Yes
T: y lo guardaste	T: And you kept it.
L: guardé, sí uno, pero tenía mucha necesidad de cortarme aparte me cuesta mucho estar acá adentro, me está costando, extraño mucho afuera y no doy más	L: I kept, yes one, but I had a huge need to cut myself apart from that it 's very costly to be inside this place, it 's costing me, I miss the outside a lot and can 't stand it any more.
T: esto que me estás diciendo que te la mandaste callada unida a la necesidad de cortarte te hace olvidar lo que hemos hablado nosotras el miércoles	T: What you 're telling me about going about your business wordlessly together with the need to cut yourself makes you forget what we talked about on Wednesday.
L: sí. Tal vez que sí	L: Yes. Maybe so.
T: ¿te acordás cómo terminamos la sesión el miércoles?	T: Do you remember how we ended the session on Wednesday?
L: sí algo me acuerdo, que me dijo que quisiera que pensara, y ahora no me acuerdo la pregunta final que me hizo pero que estuve hablando de mi papá pero la última pregunta no me acuerdo muy bien	L: Yes I do remember something, that you told me you wished I would think, and now I don 't remember the last question you asked me but that I was talking about my dad but the last question I don 't remember very well.

7. References

Hamish Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. 2002. GATE: A framework and graphical development environment for robust NLP tools and applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, Philadelphia, USA, Jul. <http://gate.ac.uk/sale/acl02/acl-main.pdf>.