Computer assisted conceptual analysis. (CACAT): the concept of Mind in The Collected Papers of C.S. Peirce.

Jean Guy Meunier.

Dominic Forest Université de Montréal.

Computer assisted reading and analysis of text (CARAT) has lately explored many variants of what has become fashionable to call " text mining" strategies .   These computer approaches have been a enduring endeavour in the European reseach tradition of the   Benzecri (1981) mathematical writings   (Lebart et Sallem 1997; Reinert 1994). In North America. they find their roots in the text matrix manipulation of information retrieval(Salton 19xx , Dumais 19xx). It is only lately that they have been explored in the more humanitiy oriented texts (Rockwell and xxx 19xx,) Unsworth 2007; Hearts 19xx ; Meunier et Forest 2005 ; Zuell 19xx) . Although these strategies are robust for large corpus they mainly operate at a macro textual level are still met with resistance by the expert readers that aim   at fine and minute conceptual analysis. We present here a computer assisted strategy for conceptual analysis based on automatic mathematical classification and annotation strategies and apply it to philosophical texts.

Conceptual analysis   is an expert   interpretation methodology for the systematic exploration of semantic and inferential properties of set of predicates expressing a particular concept in a text or in a discourse. (Ayer, 1973 ; Desclés, 1997 ; Fodor, 1998 ; Brandom, 1994 ; Gardenfors, 2000 , Rastier, 2005).CARAT is the computer assistance of this conceptual analysis.

The strategy of CACAT

Our strategy rests on the following main hypothesis.

> *The expression of a canonical concept in a texte presents linguistics regularities some of which can identified through classification   algorithms*

This hypothesis itself unwraps into three sub hypothesis.

> *Hypothesis 1 : conceptual analysis can be realized by the contextual exploration of the cananical forms of a concept.*

This is realized through the classical concordance strategy and variances on a pivotal term and its linguistic variants. (vg mind, mental , mentally etc) (Pincemin et al. 2006McCarthy, 2004 ; Rockwell, 2003).

> *Hypothesis II The exploration of the contexts of a contexts of a concepts is itself realized through some mathematical classification strategy.*

This second hypothesis postulates that contexts of a concept present regularities that can be identified by mathematical clustering techniques that rest upon similarities found

among contextual segments. (Jain *et al*. 1999 ; Manning et Schütze, 1999;Manning Prabhabar et Schütze, 2007; Baeza-Yates et Ribeiro-Neto, 1999, p. 438).

> *Hypothesis IIIClasses of conceptual of similar conceptual contexts can be annoted so as to categorized their semantic content.*

This last hypothesis allows to associate to each segment of a class of contexts some formal description of their content be it semantic , logical, pragmatic , rhetorical etc. (Rastier et al. 2005 ; Djioua Brahim et Desclés, 2007 ; Meyers, 2005 ; Palmer *et al.*, 2005 ; Teich *et al.,* 2006).. Some of these annotations can be realized through algorithms others can only be done manually.

Experiment.

From these three hypothesis emerges a experiment which unwraps in five phases and is applied to the *Collected Papers of CS Peirce*. The volume I-VIII. The analysis was applied to the concept of MIND in this corpus.

Phase 1 : Text preparation

The first phase is the   preparation of the text for admissible analysis . Various operations of selection , cleaning, tokenisation, segmentation are applied . But no lemmatisation was used. The corpus so prepared gave 74450 words ,with a lexicon of 2831 word types.

Phase II

A classical concordance is made from the on the pivotal word MIN   This produced a 1798 contextual segments of an average of 7 lines each. Being to huge for a human reading, a random sampling of 1 out 2 was applied which delivered 717 contextual segments. And composed of 3071 words (tokens) and 1527 type words.

Phase III

This concordance is in itself a subtext . A classical clustering1 technique was applied to it. It generated 83 regrouping with a mean 8.3 segment per class. It is possible to represent spatially the set of words in each class . Figure X illustrates such a regrouping for the cluster no 1 .

---

[1] Plusieurs fonctions de classifications sont applicables On peut penser ici aux techniques de Clustering, d'analyse factorielle d'analyse en composante principale, aux réseaux de neurones, aux KMeans aux algorithmes génétiques, etc. Le choix.de l'une ou l'autre méthode de classification dépend de plusieurs facteurs et paramètres. (Sebastiani, 2002 ;)
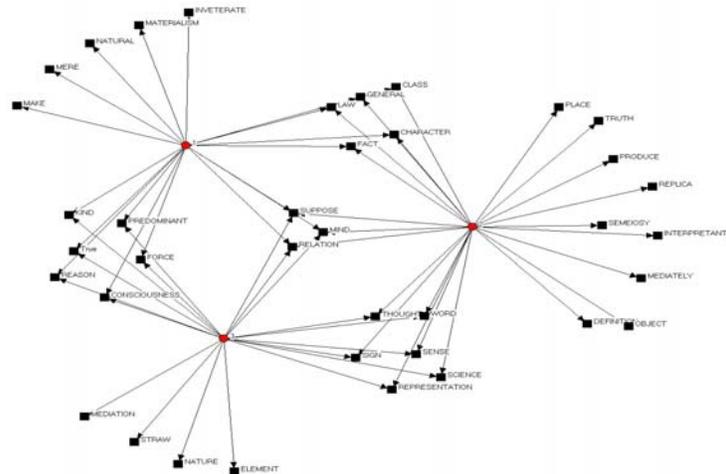
It is often on this type of representation from were   many numerical analysis start their interpretation. . One classical critic presented by expert analylists is their great generality and ambiguity . They give hints on the content but as such it is difficult to use for fine grained conceptual analysis. It must hence be refined . It is here were the annotation phase comes into play.

Phase IV annotation.

The annotation phase allow the expert reader   to make more explicit the type of information contained in the segment. For instance, the interpreter may indicate if it is a THEME, a DEFINITION, a DESCRIPTION an EXPLANATION, a ILLUSTRATION, an INFERENCE, or what is it MODALITY: epistemic, epistemological, deontic , etc. The variety of these types of annotation is in itself a research object and depend on various text and linguistic theories.

In this experiment , the annotation were realized « manually » . Although , this may take time, expert readers appreciate this for this allows them   a strong control on   the interpretation. One must remember that CARAT must assist the interpretation and not substitute algorithms to the process.

Annotation results.

The present summary does not allows us here to present the detailled results of classification and annotation. We shall only present a sample on a few segments of three classes .

Annotations on cluster no 1 :

The first cluster contained 17 segments all of which have received an annotation. Here are samples of annotation for two segment of cluster 1. The annotation is preceded by the citation itself from the original text.

[SEGMENT NO : 512]

« Finally laws of mind divide themselves into laws of the universal action of mind and laws of kinds of psychical manifestation. »_

ANNOTATION : DEFINITION : the law of mind is a general action of the mind and a psychological manifestation

[SEGMENT NO : 1457]

But it differs essentially from materialism, in that, instead of supposing mind to be governed by blind mechanical law, it supposes the one original law to be the recognized law of mind, the law of association, of which the laws of matter are regarded as mere special results.

ANNOTATION : EXPLICATION : The law of mind is not a mechanical materialism.

Annotations on cluster no 2

The same process of annotation is applied in cluster no 2 and 3 . Here is are a few sample of this annotation**.**

[SEGMENT *no: 646*]

[My definition of a sign is:] A Sign is a Cognizable that, on the one hand, is so determined (i.e., specialized, bestimmt,) by something other than itself, called its Object,?4 while, on the other hand, it so determines some actual or potential   Mind   ", the determination whereof I term the Interpretant created by the Sign, that that Interpreting Mind is therein determined mediately by the Object.?5_Peirce: CP 8.178 Cross-Ref:??_   178."

ANNOTATION   définition of a Sign : its object

[SEGMENT *no: 1033*]

"Moreover, signs require at least two Quasi-minds; a Quasi-utterer and a Quasi-interpreter; and although these two are at one (i.e., are one"   mind   ) in   the   sign   itself,   they   must nevertheless be distinct:
ANNOTATION: Explanation Sign and quasiminds:

Annotations on cluster No 3 :

[SEGMENT *no: 100*]

"But consciousness, for the reason just stated, is not to be so reinstated without tychism; nor can the work be accomplished by assigning to the"   mind   an occult power, as in two theories to be considered in the section following this.

ANNOTATION *Explanation : Consciousness is not a occult power.*

[SEGMENT *no: 279*]

" Almost all the psychologists still tell us that   mind   is consciousness".

ANNOTATION*: DEFINITION :*   Mind is consciousness (in psychology))

Phase V The interpretation operation.

The last phase is the interpretative rereading of the annotations. Here the interpreter situates the annoted segments into his own interpretative world. He may regroup the various types of annotation DEFINITIONS, the EXPLANTION etc. and hence build a specific personal data structure on what he has annoted. From then on, he may rephrases these in his own language and style but most of all   situate them in some theoretical,

historical, analytical, hermeneutic, epistemological, etc horizons . It is the moment were the interpreter offer is synthesis of the structure he believes underlies the concept.

We present here a sample of the synthesis of conceptual analysis assisted by the CARAT process on the clusters: Here is an example of such a interpretation on the first cluster.

The concept of MIND in CS Peirce:

In cluster 1 :

The law of Mind : association

The Peircian theory of MIND postulate that a MIND is governed by law

"One of these law- a fundamental one, is associative (segment 512) . This law descrives a habitus acquires the Mind when it functions. (Seg : 436)

Association is connectivity :

This functioning est one of relation bildging by connection ((seg 507) . The connectivity is of specific nature. It realized a synthesis (à la Kant) It is a form of " intellectual" generalisation (seg: 507)

Physically realized

Such a law is also to be found in the biological world . A law that can be understood as accommodation (seg : 1436) In fact this law is the specific form of the Mind's dynamic. It is a fundamental law . But it is not easy to observed because were are victim of a interpretative tradition (seg 1330) that understand the laws of mind as laws of nature. This is a typical characteristic of an " objective idealism" (seg 1762, 1382) .The law of mind do not belong to a mechanist materialism (seg 90, 1382)

Variety of categories

The exist subdivisions of this law. (375, 325) They are related to the generalisation process that is realised in infanthood, education, and experience. It is intimately related to the growth of consciousness.

In cluster 2 and 3

The same synthesis is applied of the second cluster: Briefly summarized Here : Mind is understood in terms of a semiotic structure:

:Mind operates with signs It is a semiotic process (646) 746) Signs have object and interpretans (178, 646,) and can be classified ((269) and related to propositions and truth . 698, 699)

In cluster 3: *Mind related to consciousness279, 291 which contains feeling , willing and knowing, 447) and attention, It is a dynamic stream 1168*

General interpretative synthesis

 Briefly sumerized : *The theory of mind in Peirce collectyed Paper   as revealed in the to the three first clusters sees the MIND as a dynamic semiotic system. It operates on signs (indexical) that are related to proposition and their truth This system is regulated by associative laws Consciousness is an attribute of MIND, It is not an entity   nor a unification. It is a stream.*

Conclusion

The preceding research explored a CARAT methodology.   The classification and annotation strategy manages to regroups systematically segments of text that present some content regularity. This allows the interpreter to focus directly on the   organized content of the concept under study . It helps reveal its various dimension : definition, illustrations, explanation, inferences etc.

Still this research is limited . More linguistic transformation should be applied so as to find synonymic expressions of a concept and various summarizing and extraction of the regularities of each class are to be explored.

But the results obtained, first reinstates the pertinence of the concordance as a tool for conceptual analysis. But it situates it in a mathematical surrounding aim at unveiling the various dimensions of a conceptual structure. Most of all, we believe that this methodology   may possibly interest expert readers and analysis   for it gives a strong handle and control on their interpretation process although assisting them throughout the process.
.

References

Cossutta, F. (1995). *Pour une analyse du discours philosophique*. Langages, 119, p.6.

Dixsaut M. (1985) Le naturel philosophe. Essai sur les dialogues de Platon. Paris , *Les belles lettres-Vrin*, p.36.

Loiseau, S. (2003). Philosophical discourse from autonomy to engagement: Deleuze commentator of Spinoza, in Fløttum et Rastier (éds), pp. 36-54.

Loper, E. and Bird, S. (2002) *Nltk : The natural language toolkit. CoRR*,cs.CL. /0205028.

Hearst Marti A., *Text-Based Intelligent Systems*, edited by Paul Jacobs, Lawrence Erlbaum Associates, 1992. in Direction-Based Text Interpretation as an Information Access Refinement Computer Science Division, 571.

Ayer, J. A. (1973) *Philosophical Analysis* Central Questions of Philosophy. Weidenfeld, London.

Baeza-Yates, R. and B. Ribeiro-Neto. *Modern Information Retrieval. Addison Wesley,1999*

Brandom, R.B. (1994). *Making it Explicit.* Cambridge : Harvard University Press

Cossuta, F. (2004) Neutralisation du point de vue statégies argumentatives dans le discours philosophique. Semen no 17 in *Argumentation et prise de position. Pratiques discursives*, R. Armossy et R. Koren (eds.) pp. 81-97.

Desclés, Jean-Pierre (2003) Représentations cognitives, schèmes prédicatifs, schèmes énonciatifs, *"Parcours énonciatifs, parcours interprétatifs, théories et applications"*, Ophrys, Paris, pp. 21-46.

--- (1997) "Schèmes, notions, predicats et termes". *Logique, discours et pensée, Mélanges offerts à*

`Jean-Blaize Grize,` Peter Lang, 9-36. 47.

Peacocke, C. (1992). *A Study of Concepts*, Cambridge, MIT Press.

PINCEMIN, B., ISSAC, F., CHANOVE, M., MATHIEU-COLAS, M. (2006). Concordanciers : thème et variations, in J.-M. VIPREY (éd.), *Proc. of JADT* 2006, pp. 773-784.

Meunier, J.G., Forest, D. et I. Biskri (2005). *Classification and Categorization in Computer Assisted Reading and Analysis of Texts*, In Lefebvre, C. Cohen, H. (dir) *Handbook on Categorization*, Elsivier.

Meyers, Adam (2005) Introduction to Frontiers in CorpusAnnotation II Pie in the Sky *Proceedings of the Workshop on Frontiers in Corpus Annotation* II: Pie in the Sky, pages 1–4, New York University Ann Arbor, June 2005.

`Fisher, D. H. & Pazzani, J. M. (1991) Computational Models of Concept Learning, in D. Fisher, H. Pazzani J., Michael, & P. Langley (eds.),` *Concept Learning : Knowledge and experience in unsupervised Learning.* `San Mateo, California : Morgan Kaufman Pu.`

`Djioua Brahim, Jean-Pierre Desclés (2007), "Indexing Documents by Discourse and Semantic Contents from Automatic Annotations of Texts",` *FLAIRS 2007, Special Track "Automatic Annotation and Information Retrieval : New Perspectives", Key West, Floride, 9-11 Mai*

`Floridi, L. (2002)` *The Blackwell Guide to the Philosophy of Computing and Information.* `Blackwell.`

`Forest, D. et Meunier, J.-G. (2000) La classification mathématique des textes : un outil d'assistance à la lecture et à l'analyse de textes philosophiques. In Rajman, M. & Chappelier, J. -C. (eds.)` *Actes des 5es Journées internationales d'Analyse statistique des Données Textuelles*, `9-11 mars 2000, EPFL, Lausanne, Suisse. Volume 1, pp. 325 à 329.`

Fodor, J. (1998) *Concepts : Where Cognitive Science Went Wrong.* Oxford : OUP.

Gardenfors, P. (2000) *Conceptual Spaces.* Cambridge (Mass.) : MIT Press.

`Jain, Anil K. M. N. M. and Flynn, P. J. (1999).` *Data Clustering : A Review.* `ACM Computing Surveys, 31(3) :264-323. http://www.acm.org/pubs/citations/journals/surveys/1999-31-3/p264-jain/`

`Kodratoff Y. (1999)` *Knowledge Discovery in Texts : A DÉFINITION, and Applications* `Proc. ISMIS'99, Warsaw, June.`

Laublet, P., Nait-Baha , L., Jackiewicz, A. et Djioua, B. (2002) Collecte d'informations textuelles sur le Web selon différents points de vue (chapitre 8), *"Interaction homme-machine et recherche d'information "*, Eds. Céline Paganelli (ed), Hermes.

Lebart, L. et Salem, A. (1994) *Statistique textuelle*. Paris : Dunod.

Loiseau, Sylvain (2003) Philosophical discourse from autonomy to engagement : Deleuze commentator of Spinoza, in Fløttum et Rastier (éds), pp. 36-54.

Loper, E. and Bird, S. (2002) *Nltk : The natural language toolkit. CoRR*,cs.CL/0205028, 2002.

McCarthy, W. (2004) *Humanities Computing,* Palgrave MacMillan Blackwell Publishers.

Meunier, J.-G., Remaki, L. et Forest, D. (2000). Use of classifiers in computer-assisted reading and analysis of text (CARAT). Actes du colloque international CISST 1999, Las Vegas, Nevada, U.S.A.

Manning, C. and Schutze, H. (1999) *Foundations of Statistical Natural Language Processing*, Cambridge Mass. : MIT Press.

Manning. C. D. Prabhakar Raghavan ,Hinrich Schütze *Introduction to Information retrieval* PressCambridge, England 2007 Cambridge Cambridge University Press

Meyers, Adam (2005) Introduction to Frontiers in CorpusAnnotation II Pie in the Sky *Proceedings of the Workshop on Frontiers in Corpus Annotation* II : Pie in the Sky, pages 1–4, New York University Ann Arbor, June 2005.

Bird Steven and Mark Liberman. *A formal framework for linguistic annotation*.Technical Report MS-CIS-99-01, Philadelphia, Pennsylvania, 1999. c2005 Association for Computational Linguistics.

Palmer, M., Kingsbury, P., Gildea, D. (2005) "The Proposition Bank : An Annotated Corpus of Semantic Roles". Computational Linguistics 31

Popping, R. (2000) Computer-assisted text analysis. London : Sage.

Ramsay, S. (2004) In praise of pattern, in *The Face of Text* – 3rd Conference of the Canadian Symposium on Text Analysis (CaSTA), 2004

Rastier, François (2005) Pour une sémantique des textes théoriques. *Revue de sémantique et de pragmatique*, 17, 2005, pp. 151-180.

Rastier, F. et al. (eds) (1995) *L'analyse thématique des données textuelles : l'exemple des sentiments*. Paris : Didier Érudition.

Rockwell, G. (2003) *What is text analysis, really? Literary and Linguistic Computing,* 18(2) : 209–219.

Sebastiani, F. (2002) Machine learning in automated text categorization. *ACM Computing Surveys,* vol. 34, no 1, p. 1-47.

Teich, E., Bateman, J. A., Eckart, R. (2006) Corpus annotation by generation Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora 2006, pages 86–93, Sydney, July 2006. c2006 Association for Computational Linguistics.

Unsworth John (2005) New Methods for Humanities Research Lyman Award Lecture 2005 manuscrit disponible sur le Web.

Valette, Mathieu (2003) Conceptualisation and Evolution of Concepts. The example of French Linguist Gustave Guillaume, in Kjersti Fløttum et François Rastier (eds) Academic Discourse— Multidisciplinary Approaches,

Unworth, J. (2006) Définition data mining All the King's Horses and All the King's Men Couldn't Do Text-Mining Across the Big Ten in The Nora Site project.