Clustering of Affective Categories in Scandinavian and Romance Languages

Gabriela-Alina Sauciuc
Center for Semiotics, Århus University (Denmark)

Abstract
Cross-cultural comparable data collected in several Romance and Scandinavian languages with the help of a series of tasks inspired by Fehr and Rusell's prototype approach to emotions are analyzed from a perspective that goes beyond the interpretation of prototype theory and aligns with the theory-based approaches to categorization (cfr. the seminal paper by Murphy and Medin 1985). Due to space limitations, only the data obtained with the help of a supplemented version of the traditional *free-listing* task will be considered here. Special attention is devoted, not as much to frequency of mention, order of mention and indexes of salience – parameters that are usually discussed in categorization and cultural domain studies – but first and foremost to the patterns of clustering and the relationships holding between the various categories mentioned by every single informant and across the lists, as made obvious by the data. Domain access point is identified as a privileged position that tends to coincide with categories promoting richest conceptual connections with other category members, while the prevailing relationships between mentioned categories appear to be metonymic, metaphoric, similarity or contrast.

Proceedings of the 10th World Congress of the International Association for Semiotic Studies (IASS/AIS)
Universidade da Coruña (España / Spain), 2012. ISBN: 978-84-9749-522-6 Pp. 829-836

1. THE FRAMEWORK

1.1. Cognitive and Semiotic

This paper is intended to provide a glimpse on how semiotics and cognitive sciences may work together into a cognitive semiotics whose purpose is to provide new insights into the process of meaning making (as well as its communication) and its underpinnings. The topic of categorization, which I am pursuing in my research through empirical data derived from several languages, is at the very heart of both semiotics and cognitive sciences. Categorization is a fundamental tool of semiosis (meaning-making is derived from world apprehension through signs), while some cognitivists tend to equate categorization and cognition (Harnad 2005). Two areas of research that look at the same object of study can only benefit from each other's advances and this paper will hopefully give a glimpse on how this can be achieved.

The title of this paper is meant to underline this cognitive and semiotic endeavor, while summarizing my commitment to an empirical approach that relies on systematic data collection across several languages in search for evidence that will allow me understand the relationships between affective phenomena (as units of personal experience), concepts (as units of cognitive processing) and the lexical items used in talking about affective experience (as units of verbal communication). Up to date, 10 different sets combining quantitative and qualitative data have been collected, corresponding to the various relevant superordinate categorizations in 6 languages, with Norwegian and the three Romance languages that are part of the sample having 2, and Danish and Swedish only 1; representatives of Slavic languages, as well as data collected in Greek will soon be added to this inventory.

1.2. Categorization Theorized – From Prototypes to a Theory-based Approach to Categorization

For more than 2 millennia, concept and categorization investigations have been dominated by the so-called «classical view», according to which membership is constrained by singly necessary and jointly sufficient features. During the '70-s however, a growing body of evidence put forward by Rosch and her colleagues pointed out that categorization might be instead a probabilistic process and thus membership relies on properties that are only characteristic or typical. Rosch argued that perceptual experience seems to naturally partition the world into categories and categories tend to form around perceptually salient stimuli. However, similarity as a basis of concept and category formation was not abandoned, but revised to fit a model inspired by semiotics and philosophy of language, importing into psychological theory Wittgenstein's notion of *cluster concepts* and the organizing principle of *family resemblance*.

The enthusiasm awaken by prototype theories and studies confirming typicality as an organizing principle of categories across several domains was soon to be tempered by experimental findings revealing the inadequacy of the similarity-based approaches in addressing effects obtained in categorization experiments. In this context, alternative proposals have been advanced independently by cognitive scientists having in common the fact that category structure and concept coherence was explained by people's theories (i.e. knowledge) of the world (cfr., for instance, the seminal paper by Murphy and Medin 1985).

2. THE METHOD

A prototype theory-based methodological paradigm[1] inspired by Fehr and Russell 1984 (Experiments 1-3) has been initially adopted to suit the purposes of my investigation. Following Rosch's model, Fehr and Russell suggested that emotion concepts are prototypically organized, and thus typicality is expected to predict probability and order of output (in a listing task), as well as probability of membership (in a semantic categorization task). The three interdependent tasks that were part of the original design (free-listing, semantic categorization and prototypicality rating) failed to obtain data complying with prototype theory predictions and thus have been extended in design and analysis in line with a theory-based approach to categorization. For example, the free listing task, from which the data presented in this paper are derived, was supplemented with a successive ranking task for establishing goodness of example, as well as with an inferential task for testing to which extent a similarity-based comparison is employed in selecting best exemplars (or prototypes). Due to space limitations, only a particular aspect that characterizes data elicited with the help of the free listing task will be addressed here.

2.1. The listing task

The listing task is a very simple form of systematic data collection frequently employed in psychological, anthropological and ethnographic studies. The instructions provide informants with a general category and ask them to list examples; informants are offered also a model on how to perform the task. The time allotted for completing the task is usually restricted to a few minutes in order to encourage spontaneous «top-of-the-cognitive deck» responses (Smith et al. 1995).

An overview of the informants participating in a free listing task, as well as of the general categories selected for eliciting free-listing data in the 6 languages covered by this study can be seen in the Table 1 below:

Table 1 – Informants overview

Language	Item	Participants				Age	Examples		Dist. items
		Total	M	F	NM		Total	Av. / informant	
Danish	FØLELSE	169	37	131	1	21,69	1646	9,76	398
Norwegian	FØLELSE	102	49	53	2	26,32	1251	12,26	377
	KJENSLE	92	43	45	4	24,63	1114	12,11	340
Swedish	KÄNSLA	98	51	45	2	24,28	986	10,06	335
Italian	SENTIMENTO	178	58	118	2	21,80	1939	10,89	325
	EMOZIONE	114	49	62	3	22,71	1446	12,68	440
Romanian	SENTIMENT	153	35	117	2	20,29	1734	11,30	348
	EMOȚIE	100	18	82	0	20,6	1156	11,73	359
Spanish	SENTIMIENTO	119	61	42	16	20,6	1833	15,40	418
	EMOCIÓN	110	40	53	17	21	1493	13,57	340

[1] Fehr and Russell's is not the only attempt of approaching the affective domain from a prototype-based perspective. An alternative paradigm, replicated in several languages, was employed by Shaver et al., 1987. An overview and assessment of these various paradigms is presented elsewhere (Sauciuc 2009, «Affective categorization in Romanian», forthcoming).

Column 4 through 7 are concerned with informant parameters (gender distribution of the population - male vs female vs unknown – and average age), while columns 9 through 10 are concerned with elicited data statistics: number of items elicited by every single category, average number of examples provided by participants, and total number of distinctive items for each category. Of these distinctive items, only 70-110 were mentioned by more than 3 informants in every single data collection session.

Traditionally, the parameters that are assessed based on free listing data, for instance in prototype-based research on categorization or cultural domain research, are frequency and order of mention. In prototype theory, the results of the task are correlated with data obtained in other tasks for confirming «prototypical internal structure». In processing the data, the focus is on frequency of mention, with some indications on order of mention, while cutoff varies randomly from one author to another. In cultural domain research, instead, the parameters attended to are both order and frequency of mention across lists and various indexes of salience are computed; sporadically, attention is given to the relations between items (cfr. Sauciuc 2008 for a comparison). The task can have a goal in itself, that of eliciting elements of a general category (or domain) and allows to assess, in terms of cognitive salience, the relative similarity and difference between elements within a domain across groups of respondents (Thompson & Juan 2006); alternatively, the free listing task is employed for collecting material for subsequent stages of investigation.

3. THE RESULTS – CLUSTERING CATEGORIES

As outlined above, a number of cognitive scientists have contested the fact that Rosch's data would directly explain category structure or uncover conceptual representation. Instead, a theory-based view was advanced maintaining that people's categories and derivation of conceptual structure are theory driven, i.e. guided and explained by people's knowledge of the world. One limitation of a prototype-based approach to categorization (or of all similarity-based approaches, for that matter) which was emphasized in these criticisms was the lack of focus on intra- and inter-category relationships. This aspect, however, it was claimed, is addressed by an alternative theory- or knowledge-based approach. Such claims, but also a phenomenon observed during data input, draw my attention to the fact that informants might not base their listing on a comparison process, thus selecting examples in an order that would reflect goodness of example. Their listings are however not random, and the unfolding of patterns of mention will be shown to be significant and nuance category organization. As an illustration, consider the two Romanian examples below:

> Romanian: *iubire – gelozie – ură – pasiune – bunăvoință – iertare - prietenie* (English: love2[2] – jealousy – hate – passion – goodwill – forgiveness – friendship)

[2] The numbering of 'love 1' and 'love2' is meant to reflect a neutral, alphabetical order; no hierarchical, frequency of use, relevance, or other ordering is intended by this. For the relationships between Romanian *dragoste* and *iubire*, cfr. Sauciuc 2007.

Romanian: *fericire – ură – minciună – infidelitate – tristețe – mânie - responsabilitate – emoție - iubire* (English: happiness – hate – lie – unfaithfulness – sadness – anger – responsibility – emotie – love 2)

The fact that the items in the examples above unfold in the form of scenarios should be nothing of a shock: affective experiences are better conceptualized as events or scenarios, and thus it is plausible that people may organize them as causal sequences, leading to something that can be interpreted as a sketchy narrative. Patterns of mention do not exploit only these causal links, even though causal reasoning has been found to be very pervasive in people's listings. As the two examples from Danish and Swedish will show, similarity or contrast relationships, as well as other metonymical relationships (for instance, physiological or behavioral concomitants of affective experiences are significantly listed as examples of affective experiences) can also be exploited:

Danish: *glad - sur - vred - skuffet - ked af det - trist - sørgmodig – lykkelig – hidsig – oprevet –bange – angst – nervøs – usikker - selvsikker – ængstelig – bekymret – utålmodig – rastløs – urolig – rolig – afslappet* (English: glad, sour, angry, disappointed, sorry, sad, sad-melancholy, happy, hot-tempered, shocked, afraid, anxious, nervous, unsure / unsafe, self-assured, uneasy, worried, impatient, restless, troubled / agitated, calm, relaxed)

Swedish: *ledsen – glädje – tårar – le – gråta – skratta – rädd – trött – ont – känslig – svartsjuka – avundsjuk – saknar - längtar* (sad – joy – tears – smile – cry – laugh – afraid – tired – hurting – sensitive – jealous – envious – missing – longing)

Frequency and order of mention can be affected also by entrenched sequences, such as in the following example from Italian. The recurrence of the triad *sole-cuore-amore* was high enough to make me inquire further into its source and find out that the lyrics of a rather popular song (Valeria Rossi's *Tre parole*, reading in its very beginning *Dammi tre parole: sole, cuore e amore* «Give me three words: sun, heart and love») would be the culprit in this case:

Italian: ***sole – cuore – amore*** *– amicizia – odio – passione – freddo –caldo – insensibile – noia – triste – cupo – piacere – sesso – felicità – malinconia – bacio – canzone* (Emglish: sun – heart – love – friendship – hate – passion – cold – warm – insensitive – boredom – sad – gloomey – pleasure – sex – happiness – melacholy – kiss – song)

There are not infrequent cases where informants are so much trapped in unfolding these patterns that their examples may go beyond the affective domain or beyond the task at hand:

Norwegian, Bokmål: *Glede – sorg - latter – gråt – kjærlighet – hat – varm – kald – lidenskap – likegyldighet – blomst –fugl –farmor – søsterbror – kusine – nevø – niese* (English: Joy-sorrow-laughter-cry-love-hate-warm-cold-passion-indifference-flower-bird-grandmother-uncle-cousine-nephew-niece)

Spanish – Castellano: *alegría – compartir – familia – hermanos – amigos – conocidos – lealtad – silencio - viento suave - aire calido – luz – sol – estrella – campo – memoria – caliente – contacto – cariño – cerca – caricia*, etc. (English: Joy - share - family - brothers - friends - acquaintances - loyalty - silence - gentle wind - hot air - light - sun - star - field - remembrance - warm - contact - affection - close - caress)

The next question to ask after noticing this patterning was whether it would amount to any statistical significance so that it would be likely to affect the values on traditional parameters and thus lead to spurious frequency and order of mention coefficients, which in turn would cast a different light onto the predicted correlation between typicality ratings and probability and order of output. The results of 2-by-2 matrix analysis looking at co-occurrences of items show that indeed this might be the case, underlining at the same time interesting cultural differences in the way affective frames unfold in a free-listing task. As summarized in table 2 below, in all 6 languages in the sample, there is a clear tendency for «hate» to be mentioned after «love», for «sadness / sorrow» to be mentioned after «joy / happiness», for «happiness» to be elicited by «joy». Categories in the cluster of «anger» appear to be elaborated only in Scandinavian languages, where they occur significantly after «joy». Romance languages, instead, exhibit elaborated clusters of AVERSION, with «hate» being likely to elicit mentions of «jealousy», «envy», «resentment» and of AFFECTION, with «love» eliciting mentions of «friendship», «care», «affection», etc.; Scandinavian languages, instead, elaborate on «love» in terms of an opposition between «being in love» and «love».

Table 2 – Ranges of co-occurrences

Language	Target category	Eliciting category	% range
All	Hate	Love	23, 81-61,54
All	Sadness / sorrow	Joy / happiness	12,50-50
All	Happiness	Joy	10,47-23,53
All Romance	Joy	Happiness	10,14-20,69
All Scandinavian	Anger	Joy	14,81-45
All Romance	Friendship, care, affection, etc.	Love	9,68-33,33
All Romance	Jealousy, Envy, Resentment	Hate	14,29-24,24

Most of the cases presented in Table 2 indicate the presence of asymmetrical, unidirectional relationships where one category tends to overwhelmingly elicit the other one possibly leading to spurious frequency and rank (order of mention) coefficients. In some cases, however, such as that of Romance joy-happiness, the relationships appear to be symmetrical, thus most likely they generate valid frequency and rank coefficients.

4. DISCUSSION

Meaning-making is the central issue of semiotics. However, the understanding of the processes, formats and contents involved in meaning-making are intimately related to categorization and conceptualization, topics which are at the heart of cognitive sciences. During the last decades,

categorization theorizing underwent several shifts, from the classical view postulating membership on the basis of necessary and sufficient features to a series of probabilistic alternatives, and from a similarity-based approach to a theory- or knowledge-based approach. This paper is intended as an illustration of a cognitive semiotic approach to categorization and brings evidence that, unlike predicted by prototype theory (or any similarity-based account for that matter), humans might not employ similarity matching in categorization-related tasks. It has been shown that listing of examples, which traditionally is claimed to be predicted by goodness of example (or typicality) of members, is not based on a comparison process and that exemplars are not accesses according to their typicality. Instead, the analysis of patterns of mention, based on data collected across 6 languages, shows that, in producing examples, informants exploit different conceptual relationships which may involve similarity-contrast, but also metonymical and metaphorical relationships, with causal reasoning being dominant.

Judging from the examples provided by the data, it appears that the first item mentioned, termed here domain access point, drives successive listing and that availability and affordability of conceptual relations between items (similarity, contrast, metonymy, metaphor) contributes as well to increasing frequency and rank coefficients of listed terms. This is not without important methodological consequences, considering that, based on the original prediction that goodness of example predicts probability and order of output, rather often, typicality of category members are simply established by looking at frequency scores in a listing task. It has been shown elsewhere (Sauciuc 2009a, Sauciuc2009b) that these correlations are not reliable. This paper brings further evidence in this direction by contemplating the possibility that frequency coefficients based on free listing may be spurious and dependent on how informants develop conceptual frames. Thus, I suggest that patterns of mention constitute a valuable, yet unexploited resource in (affective) categorization research and that a systematic analysis of these patterns is needed before ascribing prototype / cognitive salience status based on traditional parameters.

The analysis of these patterns of mention indicate, as well, that the listing task is not fully decontextualized and that there might be at least three kinds of contextual factors affecting the measures. On one hand, there is the personal context of the informant which becomes a source of bias only in very specific conditions (for instance, when large groups of the sampled population are involved in activities with strong affective impact). The general category which is employed for eliciting data may also influence the strategies employed by informants in providing their examples. And finally and most important, informants are retrieving context during listing by exploiting theoretical knowledge associated with listed items (in the sense of Murphy and Medin, 1985).

REFERENCES

Fehr, Beverly, Russell, James A. (1984): «Concept of Emotion viewed from a prototypical perspective», *Journal of Experimental Psychology – General*, 113: 464-486.

Harnad, Stevan (2005): «To Cognize is to Categorize – Cognition is Categorization», H. Cohen, C. Lefebrve (eds.), *Handbook of Categorization in Cognitive Science*, Elsevier, Amsterdam etc., 20-43.

Murphy, Gregory L., Medin Douglas L. (1985): «The Role of Theories in Conceptual Structure». *Psychological Review*, 92(3): 289-316.

Rosch, Eleanor (1975): «Cognitive Representations of Semantic Categories», *Journal of Experimental Psychology – General,* 104(3): 192-233.

— (1999[1978]): «Principles of categorization» in eds. E. Margolis, S. Laurence, *Concepts: Core readings*, Cambridge, MA: MIT Press, 189-206.

Sauciuc, Gabriela-Alina (2007): «*DRAGOSTE and IUBIRE* – Love in the layperson's model of native speakers of Romanian», A. Cunita, C. Lupu, L. Tasmowski (eds.), *Studii de Lingvistica si Filologie Romanica – hommages offerts a Sanda Reinheimer-Ripeanu*, Bucharest, University of Bucharest Press, 484-517.

— (2008): «What free-lists tell about categorization – The case of emotions» – paper presented at the meeting of the Nordic Research School in Linguistic, November 2008.

— (2009a): «Affective categorization – Some methodological concerns» – paper presented at the 2nd SALC (Scandinavian Association for Language and Cognition) Conference, June 2009, Stockholm, Sweden.

— (2009b): «Laymen's concepts of affectivity – general categories in 6 languages» – paper presented at the ISRE (International Society for Emotion Research) Conference, Leuven, Belgium, August 2009.

Smith, Jerome J., et al. (1995): «Salience Counts: A Domain Analysis of English Color Terms», *Journal of Linguistic Anthropology,* 5(2): 203-2.

Smith, Jerome J., Borgatti, Stephen P. (1997): «Salience Counts—And So Does Accuracy: Correcting and Updating a Measure for Free-List-Item Salience», *Journal of Linguistic Anthropology,* 7(2): 208-209.

Thompson, Eric C., Juan, Zhang (2006) «Comparative Cultural Salience: Measures Using Free-List Data», *Field Methods,* 18 (4): 398-412.