Inference and Categorization

Françoise Gayral, Daniel Kayser

LIPN
Avenue J-B. Clément 93430F- Villetaneuse
{fg,dk}@lipn.univ-paris13.fr

Abstract

In the general framework of natural language sentence understanding, categorization appears as useful for two main purposes: (1) categories are needed in order to account for the ability to cope with a virtually infinite set of sentences, (2) understanding implies the ability to infer; now, drawing the appropriate conclusions cannot be derived on a case-by-case basis but on the basis of some categories.

We propose to address simultaneously these two issues in ba s-ing categories on inference invariance. Linguistic units are grouped together as far as they share common features in their inferential behavior. Our viewpoint differs from classical approaches in at least four respects: (1) the criteria used for categorizing, (2) the fact that the categories we are looking for are neither universal, nor identical to human ones: they only correspond to groupings leading to correct inferences with respect to a specific context, (3) the status of categorization: it is not a goal in itself, but merely a means to reach correct conclusions, (4) the acknowledgement of some kind of circularity in the process of category combination.

The solution we propose is based on a non-monotonic logic: the defeasibility of its inferences leads to define admissible conclusions in terms of a fixpoint equation, and this gives a correct account of the issue of circularity.

Introduction

Categorization is usually defined as the ability to classify, i.e. to group together in a category those things that look similar, and to decide which objects "fall under" a given category.

This aspect which essentially concerns "simple" categories, often named by simple words, has been the most extensively studied. This approach to categorization meets some difficulties as soon as complex categories, corresponding to more than one-word expressions, are considered (Rips 95) (Osherson Smith 81) (Hampton 95). Whatever the framework chosen (prototypical or extensional), even simple combinations such as *adjective noun* or *noun noun* are problematic: their comprehension is not strictly compositional and requires elaboration, often on the basis of world knowledge (Murphy 88).

Now, these questions are crucial for language comprehension, which must deal with complete sentences. In this area of research, more than the ability to classify and to combine, the ability to reason, to infer is fundamental.

In this paper, we attempt to bring new insights to the categorization issue, in considering that categorizing is not a goal in itself but must be driven by specific applications: for us, sentence understanding. In this framework, categories allow to account for the ability on one hand to cope with a virtually infinite set of sentences and on the other hand to draw appropriate conclusions. Addressing simultaneously these two issues presents a number of differences with respect to classical studies in this area.

First, the categories we are looking for are neither universal, nor identical to human ones. Just designed for a given purpose, they only correspond to groupings leading to correct inferences with respect to a specific context. The criterion we use to group together units is inference invariance: "units" will belong to the same category as far as they share common features in their inferential behavior. This point will be developed in part 1.

Second, once the categories have been defined for a specific context, assigning a unit to a category is not a goal in itself, but merely a means to reach correct conclusions; so, assignments must not be considered as fixed once and for all: they can change during the reasoning process, depending on the combination with other units present in the context, or on new information brought by the unfolding text. Many examples in part 2 will illustrate this point.

Finally, the process of finding an appropriate category for an expression cannot be considered as purely compositional. Some kind of circularity is clearly at play here: the category c of a word depends on the categories c_1, … c_n of the words of its co-text, which in turn depend on c. This circularity can be solved in considering this process as reaching equilibrium when all the constraints have been accounted for. The framework we propose in the last part is based on a non-monotonic logic, which defines this equilibrium in terms of a fix-point equation.

1. Objectives of the Categorization Process

As said above, we do not consider that there exists a unique and universal set of categories, which would be relevant for every task. Even if we do not deny the existence of stable categories which are parts of our conceptual repertoire, some experiments have shown that people, in their every day cognitive activities, are able to form ad-hoc categories (e.g. 'objects that prevent the door from closing', … (Barsalou 87)).

Even concrete objects of the real world, lexicalized by a single word, are tied up with different categories, according to properties that are salient in context. An egg, for example, will be categorized as an alimentary product (with meat, etc.) in the context of cooking recipes, or as a fragile object (with glasses, vases, etc) in the context of transportation.

Rather than considering a single set of categories including, from the very start, all the available information, introducing different levels for each category seems preferable. As a matter of fact, reasoning at a simple level with "crude" categories as long as it is adequate is certainly more efficient. This does not prevent to refine the categories, to get more specific inferences, but only when the need arises, following the idea of variable ontology developed in (Kayser 88).

However, at a given level, the choice of categories is constrained: they must be discriminating in order to yield exactly the expected inferences while achieving substantial savings in the descriptions.

1.1. Inferential Behavior

Let us illustrate now what we mean by "inferential behavior" in considering two sentences in which the word *exam* occurs in a school / university context.

(1) Paul left the algorithm exam on his desk.

(2) Paul takes the exam tomorrow.

Clearly, from such sentences, all humans of a given culture draw more or less the same set of conclusions. Among these, from (1):

(1a) *The subject of the exam has been written, presumably on paper.*

(1b) *The paper is now on Paul's desk.*

(1c) *Anyone having access to the desk can read it*, …

From (2):

(2a) *There will be tomorrow an examination event with one or more examiners and Paul as a candidate.*

(2b) *Paul presumably received a notification with the location and time of the event.*

(2c) *Shortly after the event, Paul will get a diagnosis of his abilities*,...

This observation leads us to speak about the inferential potential of the sentence (1) and (2).

Linguistic knowledge is clearly not sufficient, and world knowledge is strongly required to produce this kind of conclusions. For (1a) and (1b), the interpretation of the verb tense is important, as well as the acknowledgement that an *exam* is constituted of a set of questions, generally written on paper. For (2), the interpretation of the adverbial temporal explains a part of (2a), while world knowledge about the typical unfolding of this type of event is necessary for (2b) and (2c).

So, we do not separate, in the inferential potential of a sentence, the inferences that can be considered as enclosed in the sentence itself from those that are derived by adding world knowledge. This is in accordance with findings of some psychological experiences which show that readers make practically no difference between these two kinds of inference (Bransford Franks 71).

Furthermore, the inferences that can be carried out from a sentence are influenced by a large variety of factors, such as the particular situation of enunciation, the communication purposes, the granularity level of the analysis… Nevertheless, we assume that they form a relatively well-defined set, relatively to a given level and a given task.

1.2. Method

This general framework being outlined, two main questions arise:

- How to exhibit the inferential behavior of linguistic elements? And how to separate among them the common factors from the discriminating ones?

- How to determine from an expression uttered in a given context, which elements play a crucial role in the inferences?

No currently existing formal tool is capable of doing that. We just open some tracks for a methodology, the validation of which would require some extensive experimentation.

The data are, on the one hand sentences and, on the other hand, conclusions. Varying the parameters of the sentences (modification in the syntactic structure, substitution of the verb / subject / object / prepositional phrases) leads to observe differences and similarities in the conclusions; when differences are noticed, we analyze the reasons explaining them. Since conducting a real experimental protocol is beyond our skills, our approach is at the moment more or less introspective; an interdisciplinary collaboration should, in a later phase, validate or invalidate our intuitions.

We illustrate this methodology on the word *exam*:

Substituting in (1) this word by *contract*, *leaflet, magazine, poem, story, discourse* does not modify the inferences (1a,b,c). All these words share with *exam* the property of containing meaningful information, possessing a written support, and the fact that the word can refer to this support. So, many relevant inferences are common and could be accessible through a category, named for example `InformationalObject`. These inferences are related to the content (e.g. *the exam / the discourse is interesting*) as well as to the written support (e.g. *the exam / the discourse was two-pages long*).

On the other hand, (2) emphasizes properties related to some sort of selective event; inferences (2a-c) remain valid if we replace *exam* by *A-level*, *driving license*, *job interview*,… all of which evoke a process by which candidates have their capacity in a particular domain evaluated by an authority. Inferences concerning the different phases of the process, its actors, some causal relations (the more a candidate has worked, the more likely he/she is to succeed) can be triggered. This entitles to define a category which can be named `SelectiveProcess`.

As said above, the decision to group different units into a category depends also on the level considered as appropriate for the task in view. The decision is taken according to the balance between the savings obtained through factorization vs. the importance of the residual specification required to define the behavior of each unit. This residue exists when:

- an unit has a behavior considered as exceptional with respect to an inferential property of its category; this behavior can be part of the normal properties of another category, or it can be specific, e.g. in the case of idioms;

- the unit determines the value of a feature more precisely than what is known at the level of the category. For illustration in the `InformationalObject` category, the nature of the content (political / union / social for *leaflet*), the frequency of publication for *newspaper* and *magazine*, the length and the style for *novel*,… are specified at the word level, in a similar way as a role specification in semantic networks (Brachman Schmolze 85).

On another side, properties can be inherited from a higher level. For example, in (2), properties such as 'occurring in some place, at a given moment',… should be factorized at a higher level than `SelectiveProcess` in a category `Process`.

1.3. Relationship with Distributional Analysis

The above method bears some similarity with Distributional Analysis (DA) (e.g. Harris 90): we would like here to stress the contrast between the two approaches. In DA, the observation of the occurrence of words in corpus is seen as a sufficient source for semantics; the category or class of a given word is equated to the set of contexts in which the word occurs: its distributions. Contrarily to our approach, a major role is given in DA to syntax, and extra-linguistic features are neglected. The same hypothesis is assumed in statistical approaches, which consider narrow contexts of the words to measure word similarity (Kohonen 77) (Lin 98). For example, in (Honkela 97), Self-Organizing Maps are used to build word-category maps that can be considered as implicit categories emerging during the learning process.

Such approaches are very helpful in the domain of Information Retrieval and confer a kind of objectivity to the process of categorization, but their final results are probabilistic in nature. In the best case, they yield a working set of categories sharing common semantic features as a by-product of the fine tuning of statistical parameters, not as a result of a principled study of legitimate inferences.

For our goal, text understanding, the latter study seems more suitable, and building categories on inference invariance massively attested under precise and controlled circumstances looks more appropriate than DA. But we should be aware that a major, unsolved difficulty about inference is the necessity to represent world knowledge and its use.

2. Context-dependent Assignment

In this section, we further constrain our scope in focusing on a specific application: the study of car-crash accident reports sent by French drivers to their insurance company. The main reason for doing so is to base our work on real texts, rather than on artificial examples. Besides, doing so limits the world knowledge at play, reduces polysemy: the word *feu*, which we discuss below, that might otherwise correspond to *fire, light, shot, burn,…* is always interpreted here as *traffic light*; finally, the situations of enunciation and the task (seeking for the responsibility of the actors in the accident) are well defined.

If we want to find relevant categories for *feu* in that context, we observe that:

♦ *Feu* as well as *stop* (both the signpost and the line on the ground), *traffic sign*, *yellow line*,… have as functionality the regulation of the traffic and are related to a "legal framework"; their meaning is pure convention; their physical materialization should be made very visible, and so on. A category `RoadTrafficObject` will bundle this information, encapsulate deontic knowledge about what this sign allows / forbids and thus trigger inferences depending on whether a user respects or not the corresponding traffic rules.

♦ *Feu* shares with *car, warning sign, engine,…* the property of being a physical object containing a device enabling the functionality for which the object has been designed. These common factors lead to propose a category `Device` with three facets: the physical object, the included mechanism, and the process related to the device.

These potential categories for *feu* being defined and the existence of general categories such as `Location`, `Time`, `PhysicalObject` being assumed, the assignment of one of these categories to a given occurrence is far from being straightforward. *The traffic light is out of order* informs on a property of the device, not of the object. Knowledge related to the category `Device`, such as '*when a device is out of order, its functionality is not available*', added to specific information related to *feu* (its functionality consists in regulating the traffic at crossroads) allows inferences such as '*the cross-roads might be dangerous*', '*traffic jams are likely in the surroundings*', …

In other sentences[1], *feu* is used as a location: in '*I stopped at the light*', '*drive up to the light*', inferences corresponding to 'where'-questions are available. But this is not specifically related to the word *feu*: a general characteristic of language is at play expressing that some words, used in appropriate contexts, receive the category of a location even though it is not their primary purpose.

A temporal interpretation is also possible, e.g. when *feu* cooccurs with *vert* (*green*): in '*while I was moving off at the green light…*' the inference corresponds to a 'when'-question.

But the previous examples show also that several categories can coexist for a single occurrence of the word. This is even more obvious in the following sentence of the corpus: in '*being stopped at the three-colored traffic light (red)*', the moment and the location are simultaneously present. More striking in this sentence is the fact that, although *three-color* and *red* seem straightforwardly belong to the same category, the two adjectives do not qualify the same facet of *feu*. The feature *three-color* qualifies the (permanent) ability of the `Device` to emit lights of three different colors, while *red* picks up a specific phase of the `Process`, and often yields inferences related to temporal and deontic knowledge. As a matter of fact, '*to pass the three-color light*' has no temporal interpretation and does not lead to con-

[1] All the examples of this section are extracted from our corpus.

clude a "wrong" behavior for the driver, contrary to '*to go through the red light*'.

These examples — as well as many others of the corpus — show that there it makes little sense to endow the word *feu* with an intrinsic category. As a matter of fact:

• if we bind it exclusively to one of the above-mentioned categories, the inferences obtained will, most of the time, be inadequate;

• if we consider it as a hybrid, say `RoadTrafficObject•Device•Location•Time`, with respect to Pustejovsky's complex types (Pustejovsky 95), we assume that the set of its possible categories can be known a priori;

• if we admit that *feu* is so special that it requires a category of its own, say `Feu`, we are sooner or later doomed to abandon the very idea of a categorization.

The only remaining solution seems thus to consider the assignment of an occurrence to a category as a defeasible choice, i.e. a choice that can be changed according to subsequent reasoning or new information. We then need to express the rules which, in a given context and enunciative situation, lead to a change in the category, knowing that a change may have repercussions over other choices. Although a universal set of such rules is an unreachable goal at present, it is not unreasonable to try and identify the rules that work for restricted domains (e.g. the uses of *feu* in car-crash reports).

3. Our Proposal

At this point, the requirements for an adequate model have to be made more precise.

♦ A symbolic approach is needed to endow the system with explicit representations of the inferential rules and of the sentences on which they operate. Categories are part of the language of representation.

♦ Most attempts to generalize meet sooner or later the problem of exceptions, and factorizing inferential behaviors into categories is no exception! Expressing a general behavior at the level of a category does not prevent from specifying some elements as being exceptional. Similarly, the rules that embody common sense or linguistic knowledge possess exceptions as well. Handling exceptions is thus crucial.

♦ The conclusions which have been drawn from the assignment of words to categories are not definitively granted but on the contrary defeasible, as new information, brought for example by the unfolding text, can lead to a better assignment. We need a model which allows to cope with this defeasibility.

♦ The context-dependence of this assignment as outlined in the previous section is sufficiently pervasive in natural language not to be considered as an imperfection. But, since it induces some kind of circularity, an approach relying on the compositionality hypothesis and thus implying a purely bottom-up mechanism would be inefficient. As an illustration, in *j'ai passé le feu* (*I went through the traffic light*), it is the mutual influence between the noun *feu* and the verb *passer* which allows to consider *feu* as a location and *passer* as a movement verb although both of them are strongly polysemic (in *passer un examen* (*take or pass an exam*) *passer* would not be spatial). We need a model that accounts for this circularity.

All these requirements motivate the need for non-monotonicity. Technically, various kinds of non-monotonic logics have been proposed, and they correspond to different formal hypotheses. Among them, the logic for default reasoning (Reiter 80) defines the notion of extension as a solution to a fix-point equation. This allows to represent the equilibrium between the interpretation of an expression and the interpretation of its co-text, and solves the circularity already mentioned.

Concretely, we use Reiter's semi-normal defaults (Reiter 80) with the notational convention: $A : B\ [R1]$ is a shorthand for $\frac{A\ :\ B \wedge R1}{B}$ which means: if A is true and if the conjunction $B \wedge R1$ is consistent with what is known, then B is true.

Consider now a word A (e.g. *feu*) which can belong to two possible categories, say B (e.g. `RoadTrafficObject`) and C (e.g. `Device`)[2]. Each possibility is enforced only if it is considered as making sense in the given context. For the sake of simplicity (a full development would, step by step, carry us too far away), we represent this state of affairs by two propositions BM and CM (read: assigning A to category B, resp. C would be meaningful in the present context).

Two defaults will be a priori applicable:

(d1) $A \wedge BM : B\ [R1]$ and (d2) $A \wedge CM : C\ [R2]$,

i.e. as long as $R1$, resp. $R2$ is consistent (see below), the meaningfulness of an assignment is a sufficient argument to consider it as a legitimate choice.

Now, in most cases, considering A as a B makes the choice of C inappropriate and vice-versa. We could express this by a "strong" statement of exclusion, e.g. $B \Rightarrow \neg C$, but this would forbid the possibility of co-presence. We adopt a milder form:

(d3) $B : \neg R2\ [R3]$ and (d4) $C : \neg R1\ [R3]$

If we want to accept the co-presence of the two interpretations, there is a possibility to override the mutual inhibition between B and C, by:

(d5) $BM \wedge CM : \neg R3\ [R1 \wedge R2]$.

To sum up, let D be the set {d1,...,d4} and D' the set {d1,...,d5} and A represent the fact that the word occurred, (i) both theories <{A}, D> and <{A}, D'> have a unique extension containing neither B nor C: this models the fact that, in the absence of any statement of meaningfulness, no interpretation of A is proposed; (ii) both theories <{A, BM}, D> and <{A, BM}, D'> have a unique extension containing B and not containing C; this models the fact that if A belonging to category B makes the sentence meaningful, then the only solution is to consider A as a B; (iii.a) theory <{A, BM, CM}, D> has two extensions (one containing B but not C, the other C but not B), i.e. in the absence of an explicit

[2] We obviously should not require the set of all possible categories to be known from the beginning of the analysis; rules such as "if attaching A either to B or C is not satisfactory, then apply rules for shifts of meaning" exist as well (Kayser Abir 1995).

permission for co-presence, if both readings are meaningful, the sentence is ambiguous; (iii.b) theory <{*A, BM, CM*}, D'> adds a third extension containing both *B* and *C*: if both assignments make sense, allowing for co-presence gives the choice between mutually exclusive readings (preserving the ambiguity), and a new reading in which the word has simultaneously the two categories.

Furthermore, the non-monotonicity takes into account the fact that the meaning of a sentence may appear later: suppose that, at the time of reading, no meaningfulness statement *BM* or *CM* is derived, hence (case i) *A* is left uninterpreted; if the subsequent lines bring evidence that the category *B* would make sense, i.e. allow the derivation of *BM*, the assignment of *A* to *B* becomes effective (case ii); conversely, if we are in case (ii) because of a default proof of *BM*, and if afterward some step of the proof is invalidated, we modify the comprehension of a part of the text, which, until then, was considered as understood.

For instance, the sentence *the traffic light is damaged* is meaningful with two exclusive interpretations (case iii.a): the pole supporting the traffic light is damaged, or the pole is OK but the lights are damaged. If we read later that *it is lying across the road*, the occurrence of *traffic light* can no longer be attached to the category `Device`, and we fall in case (ii) where only its attachment to `PhysicalObject` is sensible.

These various possibilities hint at the adequacy of semi-normal defaults to handle defeasible categories assignments, and prove this tool to be flexible enough to give a local permission or interdiction of co-presence, yielding either ambiguous readings (multiple extensions), or multiple meanings of a word as part of the same reading.

4. Conclusion

We have tried in this paper to promote the idea that categorizing is not a goal in itself, but a means to achieve savings in the descriptions for a specific goal. As our area of research is natural language processing and as in this framework, the notion of inference is fundamental, we have proposed to set this notion at the heart of the categorization system. A collaboration with other cognitive fields, e.g. Psychology to set real experiments, would be interesting to validate the initial categories.

Due to the nature of human language (and cognition), it is unreasonable to hope for an ideal categorization, bringing savings in the description of inference without counterpart. The counterpart is the necessity of handling exceptions, and this gives at least theoretically a criterion for the utility of introducing a new category: it is useful as long as it saves more description than it adds specificities to describe.

We have also shown that the assignment of a category to an expression is context-dependent, and thus liable to change during the interpretive process. This process constructs one (or more) solution(s) to a system of constraints; some of the constraints originate in the language, but other ones come from knowledge about the world and the enunciative situation. The proposed framework uses Reiter's semi-normal default logic, which provides an adequate technical toolbox to represent the equilibrium reached by a system of constraints as a solution to a fixpoint equation.

Expressing the constraints remains a difficult task that can be done only in a very restricted situation, restriction that is also needed to keep the system tractable.

References

Barsalou, L.W. 1987. The instability of Graded Structure: Implications for the Nature of Concepts, in U. Neisser (ed.) *Concepts and Conceptual Development*, Cambridge Univ. Press.

Brachman, R.J. Schmolze, J.G. 1985. An Overview of the KL-ONE Representation System *Cognitive Science* 9: 171-216.

Bransford, J.D. Franks, J.J. 1971. The abstraction of linguistic ideas *Cognitive Psychology* 2: 331-350.

Gayral, F. Kayser, D. Lévy, F. 1997. Quelle est la couleur du feu rouge du Boulevard Henri IV ?, *Verbum* 19 (1-2): 177-200.

Gayral, F. Kayser, D. 2001. Categorisation seen as factorisation of the inferential potential: application to the comprehension of utterances, *Cognitive Systems* 5(4): 345-371.

Hampton, J.A. 1995. Testing Prototype Theory of Concepts, *Journal of Memory and Language* 34: 686-708.

Harris, Z. 1990. La genèse de l'analyse des transformations et de la métalangue, *Langages* 99: 9-20.

Honkela, T. 1997. Self-Organizing Maps in Natural Language Processing, Ph.D. thesis, Helsinki.

Kayser, D. 1998. What Kind of Thing is a Concept. *Computational Intelligence* 4(2): 158-165.

Kayser, D. 1994. What kind of models do we need for the simulation of understanding?, in C.Fuchs & B.Victorri, (eds.), *Continuity in Linguistic Semantics*: 111-126, John Benjamins.

Kayser, D. Abir H. 1995. A Non-monotonic Approach to Lexical Semantics in P.Saint-Dizier & E.Viegas, (eds.) *Computational Lexical Semantics*: 303-318, Cambridge University Press.

Kohonen, T. 1995. *Self-Organizing Maps,* Springer.

Lin, D. 1998. Automatic Retrieval and Clustering of Similar Words, *COLING ACL* vol.II: 768-774

Murphy, G.L. 1988. Comprehending Complex Concepts, *Cognitive Science* 12: 529-562.

Nunberg, G. 1979. The non-uniqueness of semantic solutions. Polysemy, *Linguistics and Philosophy* 3: 143-184.

Osherson, D.N. Smith, E. 1981. On the adequacy of prototype theory as a theory of concepts*, Cognition* 9: 35-58.

Pustejovsky, J. 1995. Linguistic constraints on type coercion, in P.Saint-Dizier & E.Viegas (ed.) *Computational Lexical Semantics*, 71-97, Cambridge Univ. Press.

Reiter, R. 1980. A Logic for Default Reasoning, *Artificial Intelligence* 13, 1-2: 81-132.

Rips, L. J. 1995. The Current Status of Research on Concept Combination, *Mind and Language* 10 (1/2): 72-104.