

THE COG: MAKING SENTENCES FROM CONCEPTS

Boris Gorbis

8484 Wilshire Blvd., Suite 800, Beverly Hills, Ca, 900211, USA

Abstract: *Language is a cognitive map of concepts. A cognitive dictionary maps our shared reality as represented by linguistic competence of all speakers. COG offers a systematic description of concepts defined by their language usage. The 'usage' is formalized and the paper describes format and rules employed in constructing this new type of lexicographic COG data-bases. A concept "entry" is a description of a "spectrum of aspects", i.e. all cognitive points-of-view from which anyone can view a given concept. Each aspect is serviced by a comprehensive list of Operators (Active, Passive and Modifiers). Finally, the paper offers two ways of indexing word-forms (and concepts): by their semantic weight and by semantic index. By providing a shared metalingual framework, COG(s) might become a useful tool in human-human and human machine interactions.*

Keywords: 'cognitive dictionary', 'aspect semantics', meaning, 'shared reality'

Introduction:

This writing is an outline of basic ideas and methodology involved in creating a cognitive dictionary - a COG. It is by no means a 'scientific' theory i.e. it does not pretend (in spite of its tone and my efforts) to state verifiable, objective truths but rather follows a task set forth in my much earlier work. Years ago I authored an article in an obscure but important Soviet linguistic journal "*Translators Notebooks*" edited by L.S. Barkhudarov. (Gorbis, 1977) which (considering that I was already living in the United States) was a miracle having being allowed to be published. The article described a field which I labeled "*generative lexicography*" and it postulated the following task: to design a dictionary that would allow anyone armed with instructions on its use to generate linguistically correct sentences. The user's knowledge of the language was not specified but it was tacitly assumed that some knowledge existed or was available. This early work led to explorations in two directions: developing Universal Grammar Metalanguage (Gorbis, 2005) and collecting data on the linguistic life of concepts. It is this latter work that excited me for many years and led to shaping of the COG which in turn led to development of ideas I collected under the rubric of '*aspect semantics*.' In its most basic form this term applies to the analysis of '*shared reality*' from a speaker's perspective.

It is an effort to describe one possible approach to the task and, I confess, a highly idiosyncratic one. For example, in order to separate the ideal, linguistic, objective and cognitive lives of a concept, a terminology is being offered that may be somewhat confusing. Instead of the "*word*" the term "*word-form*" is used not only to emphasize that a word is a linguistic fact but to strip it completely of any psycho-semantic overtones in the discussion that follows. On the other hand, any reference to "*concept*" is to a purely ideal (psychological) phenomenon with total disregard of its linguistic or psycholinguistic existence. Next, we introduce such terms as "*aspect*", "*meaning*" and "*property*" as essentially equivalent notions but each dedicated to its own sphere: cognitive, linguistic, and psychological respectively. These and other terms will be defined below but first I need to describe the subject of this effort.

Dictionary As a Cognitive Map of Shared Reality

There are four types of dictionaries in the world. The best known are logical dictionaries in which concepts are described or defined by reference to other concepts. The second type is a psychological dictionary, such as a *Thesaurus* in which concepts are illuminated by their associations with other concepts. The third type is a bilingual dictionary in which a concept is referenced by equivalent words in two different languages. The fourth type is a *cognitive dictionary* in which a concept is described or defined by its usage in a certain language or pair of languages. While many lexicographical works employ usage as a tool to refine a meaning or a shade of a meaning, I know of no efforts to create a systematic description of concepts solely by reference to their usages in a given language.

One possible reason why this work was not undertaken is the misperception of the sheer magnitude of usage options available to a competent speaker. However, the “*usage*” of a concept is not an open-ended set. In fact it is a fairly small one. The other reason why concepts did not receive this treatment to date is their composition. A concept is first and foremost a psychological phenomenon, an ideal generalization. On the other hand, a concept (as represented in lexicography) is a concrete element, a linguistic event. Furthermore, a concept is an integral part of what we call a “*shared reality*” – objective i.e. real objects/states/processes we observe, experience, and manipulate in daily life. What is remarkable is the ease with which we shift gears from *psychological* to *cognitive* to *objective* to *linguistic* direction. It comes from the cognitive ability of a language speaker, from his or her linguistic competence, but we still lack a working protocol of describing this multidimensional phenomenon. Thus, our cognitive dictionary is offered as a model of that competence, i.e. a model of how shared reality knowledge is encoded, structured, accessed, and actualized by language speakers.

For starters, it is necessary to identify our challenges and then to formulate basic ideas underlying this approach and methods offered here. The challenges are many. First, it would be totally useless to simply enumerate all occurrences of a given word-form usage as some strings of text. Thus, we need to formalize what constitutes “*usage*” in a cognitive dictionary format. Next, we need to somehow group instances of usage as dictionary elements. Neither the alphabetic listing in logical dictionaries, nor the proximity criteria utilized in thesauruses are suitable for obvious reasons. The first one is psychologically meaningless and the second cannot be applied to linguistic facts. Finally, there is the challenge of completeness. Having compiled a number of formalized usages for an entry and having decided on their grouping, we still do not know what is left out.

To address these issues we introduce some fundamental assumptions and opinions underlying this lexicographical experiment. We already mentioned the notion of ‘shared reality’ which is a cognitive phenomenon. It exists essentially outside of any language and can be imagined as *a sum total of knowledge of all concepts and their instrumental and qualitative characteristics shared potentially by all humans, regardless of when or where they happen to communicate*. The “shared” here is the opposite of the least common denominator. It is additive, which means “potentially available” to any person. Shared reality is a universal cognitive phenomenon and as such it is constantly expanding and developing. Just consider the changes that the concept of “memory” underwent in the past 50 some years not only in the knowledge that “memory” can exist outside of human mind but, most importantly, in the development of means and skills (verbal and non-verbal) of creating, and using ‘memory’ as an outside device. Such knowledge would necessarily include means of instructing everyone on what a “memory” can do and what can be done to different types of “memory.” The ease with which we can now “buy memory”,

“install memory” etc. suggests the existence of a fairly simple mechanism of incorporating radically new elements into our existing cognitive structure.

Any language services shared reality by imposing a simple and thus discernable grid on the context of knowledge. Indeed, *language is a cognitive map of concepts*. Every language does it in its own peculiar way reflecting in part the environmental and in part the phonetic and morphological history of a group of speakers. (Whorf). However, because it operates from the common base of shared reality, a natural language arms its users with means of reading a potentially universal cognitive map of any concept. This map is not idiosyncratic; it does not depend on morality or culture. The cognitive map of the concept of “God” is the same for atheists in Italy and believers in China.

The ideal “cognitive map” includes potentially all that can be done to an object of our thought (concept) as well as everything that this object can do and all available criteria by which one such object (concept) can be likened to, grouped with or differentiated from other objects of thought (concepts). Our cognitive-linguistic competence of “*memory*”, “*God*” or “*hammer*” includes knowledge of what can be done with *memory*, what *God* can do and what kind of *hammers* we need or use. Thus our initial quandary as to what is “usage” and how to formalize it can be answered by agreeing that (with regard to any concept) our dictionary should have three basic elements: (1) What can be done with or to the objects represented by a given concept, (2) What can these objects do, and (3) What kinds or types of objects are there.

Our object of thought (concept) can be viewed from different angles and thus the cognitive map of the concept “*drug*” should contain clear means of differentiating not only between “*drug*” as medicine and “*drug*” as an illegal substance, between “*drug*” as a manufactured object and “*drug*” as a vegetation but also between “*drug*” viewed from the standpoint of its effects and “*drug*” viewed as an element of commerce. We call these points of view “*Cognitive Aspects*” or “*Aspects*” for short. *Our fundamental proposition is that any concept is representable as a sum total of all angles or points of view from which it can be viewed by a speaker*. Concept seen as a system embracing all possible aspects is the basic element of the cognitive dictionary. The importance of this proposition cannot be overstressed and a significant portion of what follows illustrates this idea.

The Structure of a Cognitive Database (COG)

“**COG©**” stands for “Cognitive Dictionary.”* It is a database of ‘*Concepts*’ described by ‘*Word-forms*’ and their ‘*Categories*’. An entry heading in the COG© is a noun ‘*word-form*’ with a corresponding value (its ‘*semantic weight*’) and index. Each *word-form* entry lists a number of ‘*Categories*.’ Each ‘*category*’ corresponds to a specific ‘*Aspect*,’ i.e. a cognitive point-of-view from which a speaker can view a given ‘*concept*’. An entry for each ‘*category*’ is divided into three parts: *Passive Operators*, *Active Operators* and *Modifiers*. Each part contains elements (units) of usage. Prefacing all entries is a list of *aspects* for all COG entries. Each *aspect* is assigned a number and a letter code. Each *concept* in the COG is thus identified in two ways: by a value which is arrived at by adding the numbers of its *aspects* (*semantic weight*) and by an index composed of codes for each *aspect* of this *concept*. Here are definitions of the basic terms and notions that are used with this approach. To further (un)confuse the reader, we shall attempt to separate psychological (cognitive) phenomena from linguistic events by using different terms depending on the respective field.

-
- At the present time, COG-related data collection and analysis continues for several decades and resulted in dictionary corpus of over 2,500 English word-forms with over a million units.

Concept (psy) – a generalized complex of ideas formed by extracting common features from specific instances and identified by at least one **word-form**.

Word-form (ling) – a unitary combination of language elements that refers a language user to at least one ‘**concept**.’

Aspect (psy) – point-of-view from which a **class of concepts** or a **single concept** may be considered by a language user for the purpose of generating or perceiving speech.

Spectrum of Aspects (Spectrum) (psy) - a volume of potentialities encompassing all points of view from which a given **concept** may be considered by a language user.

Cognitive-linguistic Category (Category)(psy/ling) – a database of encyclopedic knowledge of a language user with regard to a specific **aspect** that contains answers to 3 (three) fundamental questions:

(1) What can a ‘**concept**’ *do* when viewed from a given **aspect**? (2) What *can be done* to a ‘**concept**’ when viewed from a given **aspect**? (3) What kinds or types of ‘**concept**’ are there within a given **aspect**?

Active Operators (ling) – a set of word-forms that have the **property** to satisfy answers to question (1);

Passive Operators (ling) – a set of word-forms that have the **property** to satisfy answers to question (2);

Qualifiers (ling) – a set of word-forms that have the **property** to satisfy answers to question (3);

Unit (ling) – an element of set (1), (2) or (3) linguistically represented in Indo-European languages as either a V+N, N+V or an A+N combination of word-forms respectively.

Property (psy/ling) – a feature common to different **concepts** viewed from a given **aspect**.

Class of concepts (psy) – all concepts that have the same **property**.

Meaning (ling) – a potential or actual product of extraction of a **property** by **unit** from volume of potentialities.

Let us highlight these definitions by reviewing one word-form traditionally seen as the same concept with different meanings. Consider the following sentences:

- (1) *Time to go to school.*
- (2) *I went to middle school in Oregon*
- (3) *I can't find the right school for my kids.*
- (4) *I can't find the school.*
- (5) *I missed school today.*
- (6) *I miss my school.*
- (7) *I struggled through high school.*
- (8) *The entire school was evacuated.*
- (9) *I need to pay for school.*
- (10) *School was good for him.*
- (11) *I could not see the school*
- (12) *School stands at the entrance to the park.*
- (13) *The school was named after Kennedy.*
- (14) *We plan to paint the school this summer.*
- (15) *Your conduct embarrassed the school.*
- (16) *There is no school after 5 p.m.*
- (17) *I have no school tomorrow.*
- (18) *I don't trust this school.*
- (19) *School occupied the whole city block.*
- (20) *I remember this small village school.*
- (21) *We have the best school of engineering in the state.*
- (22) *Which law school did you say you graduated from?*
- (23) *We have good schools here in Oregon*
- (24) *We need to build more schools.*
- (25) *American schools cannot be compared to German institutions.*
- (26) *Schools are becoming less and less affordable.*
- (27) *There are different schools of thoughts on this subject.*
- (28) *Our schools demand better teachers.*
- (29) *I hated Sunday school.*
- (30) *School lost its accreditation.*
- (31) *School lost its best teacher.*
- (32) *School looked the same.*
- (33) *I cannot confirm that X belongs to a Marxist school.*
- (34) *We went to different schools.*

It is obvious that the concept of 'School' is examined above from many different points of view. Any number of utterances can be added but all cognitive points from which "school" can be viewed would remain fixed and few. Here is the *Spectrum of Aspects* of the word-form "school" as derived from these texts:

OBJECT

Physical Object
Man-made structure

INSTITUTION

Procedures
Tradition
Theory
Content

ESTABLISHMENT

Religious establishment
Educational establishment
Commercial establishment

HUMAN GROUP

Network
Relationship(s)

PROCESS

Time Period
Event

RESULT

Quality
Value

As a 'man-made structure' any "school" can be *built, destroyed, designed, seen, planned, approved, observed, ran through, searched, vandalized, bombed, located covered with (what?), flooded, cleaned up, repaired, painted, re-built, broken into...* etc. It can be *big, beautiful, cozy, small, tiny, huge, modern-looking, brick, two-story, old, new, spacious, ugly*, etc. Viewed from this *Aspect* a school can *stand, rise, lay in (what?) extend, cover (how many units?), contain (what?), spread, etc.*

As an 'Institution' any "school" can be *dedicated, ran by (whom or what?), ran (how?) licensed, named after (whom?), accredited, inspected, approved, accused of (what?), headed, criticized, credited with (what?), feared, restructured, loved, created in (what year?), committed to (what?), etc.* Like any other institution schools can *hire, fire, attract (whom or what?)* and be *big, small, new, old, venerable, respected etc.*

As a 'Commercial Establishment' schools can *charge (how much?), provide (what?), earn (how much?), gross (how much?), cost (how much?), budget, spend (how much?) on (what?), go out of business, profit*, and so on. The concept seen in this aspect will have Passive Operator defined units such as: *to tax a school, to license a school, to bankrupt a school, to sell it, buy it, mortgage it* and even *run it into the ground*. In this aspect schools will be *private, public, profitable, unprofitable, expensive, inexpensive, affordable, and* so on. We could go on illustrating the word-form 'school' when viewed from the above listed points of view, i.e. *Aspects*, but this is really beyond the point.**

***) To break the monotony of this discussion, let us note in passing that there are many aspects that 'school' *does not* belong to, such as 'measured quantity' 'liquid object' 'living being', 'natural phenomenon' and so on. It means that normatively none of the categories servicing these aspects would make their units available to define a meaning of the word-form "school." The key word here is *normatively*, for word-smiths and jokers, poets and writers would always try to fit new clothes on an old concept, sprinkling their speech with metaphors and similes. You may not necessarily like the "school evaporates" and "school died" but give them a chance and they will show up in some context as did "rumors kept the school boiling" and "school was overflowing with kids."

What is important is that when invoked, each *Aspect* marshals its cognitive semantic arsenal – a *Category* with units dedicated to servicing that particular *Aspect*. The preceding is less evident: somehow we recognized or identified the *Aspect* in which “school” word-form appeared.

Obviously, there is more than one concept that can be viewed from a given point of view and we called it a **class of concepts**. Thus, anything that can be viewed as a ‘*Structure*’ *Category* will have a shared array of units (*Operators* and *Modifiers*) that will bring out this *Aspect*. In this sense, a palace, a theater, a hut, a Taj Mahal, a school, a tree-house, a church, a hospital, a City Hall and a privy would have the same meaning - *Structure*. Whatever can be done **by, upon or to** the *Structure* can be done **to** any element of this class and whatever *Structure* can do, each element in the class can do. Because the word-form “school” could be examined as ‘*Time Period*’, it would share that property with such word-forms as “play,” “lesson,” “sex,” “life,” “period,” “winter” etc, thus forming a psychological class of concepts, ” and would be serviced by the same *Operators* and *Modifiers* as other concepts in the “*Time Period*” concept class, cf. such *Active Operators* as: ‘lasts,’ ‘ends,’ ‘goes on for (how long?),’ ‘goes on for (how many time units?),’ ‘continues,’ ‘lasts for (how many time units?),’ ‘takes (what?),’ ‘drags on,’ and others. Each concept in the ‘*Time Period*’ class would be served by shared *Passive Operators* such as: “to time a play, a lesson, show, activity, sleep (but not school, winter or season), “to extend a lesson, school, program, show, etc until (when?) or by (how many time units?),” “to overstay school, lesson, show, period, etc by (how many time units?)”, “to drag out a school, a play, a game, a shift, a period, a session, a lesson, etc)” “to shorten school, a play, a lesson, a show, a session, a season...),” “to end,” “to terminate,” “to speed up a lesson, a show a play, a session” as would “to shorten a play, lesson, show, period, etc” which might be included as well. The opposite is also true and because of the shared *Aspect* all *Operators*, *Active* and *Passive* that belong to the *cognitive-semantic Category* of ‘*Time Period*’ will be available to operate on each and every element of this *class of concepts*.

This may be a good time to consider an interesting nuance; some of the units listed above for illustrating the ‘*Time Period*’ *Aspect* would be repeated when we look at the “school” (and most other concepts of this class) as a ‘*Process*’ which raises the issue of differentiating the same linguistic units that service different cognitive *Aspects*. For example, is “ending school” a “time period” unit or a “process” unit? And what about the unit “restoring school”? Is it a “structure” aspect unit or a “quality/value” one? There are means of disambiguation in most languages and cultures, but this issue is beyond the scope of the present discussion. However, it does lead us to two related questions.

The first one is: where do we get ‘*aspects*’ from? The second question is related: what comes first, the *Aspect* or the *Category* which services it? (Linguistically, *Category* is a database of *Operators* and *Modifiers* servicing an *Aspect*.) The answers depend on methodology for obtaining data. If we proceed from some mental effort to compile a list of all *aspects*, we are going in the wrong direction. In the absence of a consistent protocol to generate all aspects of all concepts, we are bound to find the process arbitrary, subjective and incomplete. On the other hand, the process of compiling all possible units for a given word-form creates its own problems: the work is tedious and the hours are awful, but...the data itself suggests which aspects it serves. No guessing and creativity – just consistency in identifying and labeling the nomenclature of aspects once the database of all *Active* and *Passive Operators* and *Modifiers* is compiled for a given *word-form*. True, the process seems to have a degree of arbitrariness to it, especially in the choice of labels for aspects. But coming up with names of aspects may only seem subjective. What we call ‘*aspects*’ are fuzzy paths along which our human brain differentiates “shared reality” into concepts (Gorbis 2005) and in that sense “*aspects*” are language independent and universal. In other words, *Spectrum of Aspects* of “school” would be the same for speakers of

different languages. Thus, to answer the second question, data comes first and it is *data that dictates how to arrive at the spectrum of aspects of a given concept*. A consistent application of the protocol is to compile all data (*Operators and Modifiers*) for the selected *word-form* and then query each unit (*operators and modifiers* alike) which *Aspect* it serves, or from which vantage point it illuminates or “excites” a given *concept*. We thus arrive at a fairly consistent list of all *aspects/meanings* for the given *concept/word-form*.

Formal Rules of Concept Mapping for COG

Let us formalize this discussion by restating the above as a set of axiomatic rules:

Rule 1. Any concept can be viewed from at least one aspect and any concept may have more than one aspect.

Rule 2. Any aspect can apply to more than one concept and any concept may share at least one aspect with another concept.

Rule 3. Any aspect correlates to and invokes a cognitive-linguistic category consisting of a finite number of operators and qualifiers that service that category.

Rule 4. Each *category* has three classes of elements called Operators and Qualifiers linguistically expressed as V+N, N+V and A+ N units that apply to all concepts that share a given *aspect*.

Rule 5. By virtue of belonging to a cognitive-linguistic category each unit possesses a specific meaning with which it defines a property of each concept.

It is easy to see how this discussion can be further formalized once we arrive by consistent enumeration at the *nomenclature of all aspects of all concepts of a given language*. This relatively small list can be represented as a finite set where each element is an ‘aspect’ which has been assigned an arbitrary index code or an arbitrary value. Then each concept/word-form that corresponds to a spectrum of aspects/meanings can be expressed either as a string of indexes (semantic index or formula of a word-form) or as a number representing the total value of the spectrum (semantic weight of a word-form.) Either way, the protocol apparently accomplishes two tasks: first, it differentiates word forms by either their formulas or by their weight and second, it permits findings of equivalency between word-forms in different languages (codes.)

Let us introduce only the *semantic weight* of a *word-form*. It represents a total of its *aspects/meanings* expressed by a single number. The number is arrived at by adding the numerical values of all Aspects of a given word-form. There is no “science” to these values; they are arbitrarily but consistently assigned to each Aspect in the nomenclature of all aspects of all concepts in a given language. Assuming that this nomenclature is described in a systemic fashion in which all word-forms have a defined “*spectrum of aspects*” we can simply add the numerical values for each *aspect* for a given word-form. The resulting number is that word-form’s weight expressed as a single value. The notion of quantification of semantics is not new and this is just one possible approach to carrying out language-based tasks outside of the natural language boundaries.

Bibliography:

- 1) Gorbis, Boris “Psycholinguistics and Generative Lexicography” in “Tetradi Perevodchika” issue 14, “Foreign Relations Publishing House” Moscow, 1977 pp103-116.
- 2) Whorf, Benjamin Lee “Language, Thought and Reality: Selected Writings of B.L. Whorf” edited by John B. Carroll, MIT Press, 1956
- 3) Gorbis, Boris “A Primitive Model of Metalanguage For Universal Grammar” Proceedings of MLMTA Conference, June 2005