Cognitive Dictionary: A Representation of Shared Reality

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 to Shared Reality

Ours is a processing world. It forces us to consume, digest and store flows of data in a variety of codes. Some remarkable skills take part in dissecting, organizing and interpreting individual elements. How do we do it? Or more precisely, what model can be offered to account for our ability to operate on a conceptual level in a beehive of languages and ideas? What mitigates the fall-out from the well known incident at the Tower of Babel? What allows a Burmese to understand an Eskimos?

The immodesty of asking these questions is self-evident. It would be outright silly to suggest that one approach yields enough illumination for the answers to appear. Yet, even a peek in that direction is worth an effort.

The starting point is a notion of "*shared reality*". The reason people can communicate, understand and affect each other is due to the existence of a hypothetical map of concepts (for lack of a better descriptor). *Shared reality* is not what any given individual develops as his knowledge base. It is a sum total of what any and all individuals can incorporate into any individual cognition. *Shared reality* is not the least common denominator. It is *additive*.

The next feature of *shared reality* is that it is *operative* by nature. 'Operative' here means purposeful, teleological. It means that this abstraction incorporates not just objects, states and processes but purposes, intended uses, consequences and effects. In this context, the purpose of communication is not to impart or deliver information – it is to affect the state or action of any other person or objects by means of codes but it is made possible only by reference to *shared* reality. Elsewhere we defined 'shared reality' as "*a sum total of all concepts and their instrumental and qualitative characteristics shared potentially by all humans regardless of when and where they happen to communicate"* (Gorbis 2006 (a))

An example might be appropriate. Consider your interpretive reaction to the following sentence: "*Living and working in California puts a terrible strain on your car"*. In any language it would bring out notions of 'money', 'ownership', 'leasing', 'roads', 'repairs', status', 'mobility', 'other drivers', and so on. Before we dismiss these as mere 'associations' shouldn't we ask ourselves from which source they appeared? And since we are making an attempt at modeling, shouldn't we ask which elements of this sentence would bring forward these associations? Linguistic analysis would not tell us and neither would traditional psychological theories of 'associations'. Linguists cannot point to anything about 'living' that connotes 'driving' and psychologists would insist that the formation of associations is essentially an ad hoc, random, individualized process.

To answer these questions we suggest rethinking our starting point: the notion of an 'element'. What is an element of language? Most would agree that it is a morpheme or a word. What is, then, an element of cognition? A thought? A concept? An idea? And before you get tired of all these questions, let us remember that cognitive processes are framed and mediated by language. Yet, as we know, they are essentially universal, that is not language dependent. Because it does not matter which language mediates our inner response to 'car', 'California' or 'working' we have no choice but to go outside the (English) language and the individual psyche to resolve this seeming paradox. However, the introduction of '*shared reality'* is an attempt to do more than simply name this universal cognitive domain. It allows us to make the next step and ponder what might be its elements and how to represent them in an accessible format.

The Metaphor of Suspended Spheres

Picture, if you can, a rotating sphere surrounded by an array of multicolored lights. As the sphere rotates, a different section of its surface is illuminated by a specific color light. Now picture hundreds or even thousands of spheres suspended in midair. Now turn off all

lights except yellow. All spheres illuminated will have two things in common – the common source of light and the common property of being yellow. Let's keep the yellow and turn on the red light in the array. It would illuminate some of the spheres that are not lit by the yellow light and some that are. Now this last grouping of spheres has four properties: two common sources of illumination and two surface areas lit up. Keep on adding lights but bear in mind that our array is large but not infinite. Soon, very soon, you will run out of lights to turn on. No matter how many spheres you can picture and how many lights are placed in one array and even regardless of the number of arrays you can imagine, you would have the following 'elements' of this imaginary gizmo: (a) a source of colored light, (b) a sphere illuminated, c) an area that is lit up, and d) the property of being colored yellow, red and so on. This is my favorite crude image of a "shared reality" and its properties.

The 'light' is a "point-of-view" from which one can view the 'spheres', that is those abstract generalizations we call "concepts". Linguistically, we know them as 'words' but for many good reasons we shall refer to them as '*word-forms*'. Take a simple word-form 'car' from our previous California example. We can look at a car from a variety of viewpoints and see it as an "Object". "Moving Object", "Hollow Object", "Enclosed Space", "Contents", "Commercial Object", "Man-made Object", "Consuming Object" , "Emotional Object" and "Functional Object". Terminologically, we call these "*Cognitive Aspects of a Concept*" or just "*aspects*". This list of *aspects* is a finite set for the *concept* and *word-form* 'car'. It is an open set i.e. it can be expanded (by virtue of 'sub-aspects) but it is nonetheless a finite set with a very small number of elements.

This set is called a '*Spectrum of aspects'* for an individual concept. Each concept and corresponding word-forms in any language would have their individual and most often unique '*spectrum'* i.e. set of all *aspects* from which this concept can be viewed by a speaker. The *spectrum* of a 'car' is not the usual lexicographic *meaning* of the 'car'. Logical dictionaries, such as Webster, and psychological dictionaries, such as Roget's Thesaurus do not account for the reality phenomena of 'loving one's car' or 'hating' it, 'getting into' it, 'leasing', 'buying' or 'hitting' it or 'using it as a weapon'. From the standpoint of '*aspect semantics*' (Gorbis, 2006(a), 2006 (b)) any 'meaning' is a function of *spectrum of aspects* for the given concept as it exists in our *shared reality*. No matter which concept we choose, it would have at any given moment (synchronistically) a very limited set of aspects (typical range of 5 to 15). In a historical perspective the set grows and changes as we discussed in another presentation, (Gorbis, 2006 (b)). Just think of the concept 'memory' in a pre-computer age and its current usage or the change of an 'Amazon' from a river to a retailing giant.

As we consider all lights illuminating a single sphere we can assume that it is uniquely defined by this particular constellation of lights just as we may assume that a concept of 'car' or 'work' is defined by its *spectrum of aspects*. For example 'work' can be represented as a spectrum of these (primary) aspects: "Object', ''Space', 'Institution', 'Time', 'Group', 'Process', 'Result' and sub-categories 'Structure', 'Human Group', 'Commercial Institution', 'Emotional Object', 'Destination' etc. Even average speakers have no problem incorporating 'work' in its various aspects into a coherent string of

utterances, cf this text: "*The challenge of work (Process) is not in getting to work (Space /Destination) early or even on time but to take your daily load with a positive attitude. If you love your work (Emotional Object) you will succeed, even if your work (Result) provides you with less than you think you deserve"*. Each of the above aspects of 'work' is an element of *shared reality* that assures your understanding of the text's message whether it is delivered in Russian, Urdu or Swahili. Here as an example is the *Spectrum of Aspects* of word-form "school" as derived from texts:

SCHOOL

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*

So far, our discussion was proceeding in a horizontal plane of reference: we take one concept at a time and analyze the various *aspects* which comprise its *cognitive spectrum*. We must return to this subject to get to the heart of our discussion, but for now we shall move to a vertical plane and look at all spheres lit by yellow or red light (discussing concepts that share the same *aspect*).

Going Vertical

It is easy to note that 'car' is not a solitary recipient of our imaginary *(red or yellow)* light. It shares any of its *aspects* with many other concepts. With some of them 'car' shares most if not all of its *aspects* (e.g. shares all with "vehicle" and "truck" but with 'motorcycle' 'car' shares all *aspects* except "Hollow Object" and "Enclosed Space"). Of course there are numerous other concepts that can also be 'lit' as 'Man-Made Objects" "Moving Objects" "Commercial Objects" and so on. Clearly, any one of these *aspects*

correlates with a long list of different concepts. It is said (Gorbis, 2006(b)) that an *aspect* is a *property* of any of the concepts that share it and we call concepts with the *same property* a '*Class of Concepts'*. "Moving Objects", for example, would apply to birds and buses, robots and hyenas, asteroids and bullets. Class of "Man-Made Object" includes bullets, trains, robots and buses, whereas hyenas, rabbits and seagulls belong to the "Living Object" Class of concepts.

There are several major Categories of aspects: Category 1 "Object", Category 2, "Activity or Process", Category 3 "State", Category 4 "Institution", Category 5 Attributes", Category 6 "Space", Category 7 "Time" and Category 8 "Collection or Group". Other sub-categories may be introduced, such as "Deity" or 'Human'. Suffice it to say that each major category is further divided into subcategories, groups, and contains *primary* and *secondary* aspects for each of the divisions. For reasons of simplicity we shall deal primarily with illustrations from Category 1 "Objects".

What is remarkable is that in every human language there exists an assortment of linguistic tools servicing each specific *aspect* of any object and therefore available for use upon and with any object (and its concept) of the Class of which this *aspect* is a property. This tool-box may be linguistically determined at the level of word-forms but its content is also a product of the 'shared reality'. Let us take a closer look at this duality: on the one hand if anything can move, it can be stopped, speed up, slow down, it can approach, move away and so on and so forth. These characteristics of a moving object are metalingual -- no language has a monopoly on them. On the other hand, each and every socio-cultural group would have verbal (and non-verbal) means by which to communicate about these characteristics usually, but not necessarily, through lexical means, i.e. word-forms. Any language reflects and services *shared reality* by imposing a simple and thus easily discernable grid on this corpus of knowledge. Indeed, while concepts form the outlines in the map of '*shared reality'*, language itself is a cognitive map of concepts. It is besides the point that each language does it in its own idiosyncratic way. The key is that any language does it through some hierarchally organized levels of abstractions which eventually culminate in the metalingual and thus universal elements of shared reality.

It is a true universality, stripped of any moral, ideological and doctrinal connotations. The cognitive map of 'God' is the same for Muslims and Catholics, agnostics and atheists. Irrespective of the language or belief system any and all of these would have at their disposal cognitive facts that deity can be 'worshipped', 'obeyed' 'disobeyed' praised'. 'hated' 'killed', 'rebelled against' 'sanctified', 'cursed' 'ignored', 'known' or 'not'. They will know that such a phenomenon as 'God' can 'exist' or 'not exist', and that it can 'kill', 'order', 'send', 'descend', 'save' and 'restore' or be and do the opposite of all of the above and a great deal more. The ability to operate with these cognitive facts does not make one any more or less of a believer in god, just a competent reader of a cognitive map.

An ideal cognitive map of any concept incorporates all that it 'can do' as well as all that 'can be done' to that concept. It must also include all available criteria by which one

object of thought can be differentiated from, likened to, or grouped with other concepts, but we are jumping slightly ahead. Consider for now that because concept *A* shares an *aspect* with concept *B* it would be redundant and uneconomical to create unique linguistic elements for *A* and *B* to express that common feature. Thus, when a concept shares an aspect with other concepts (forming a Class) all of the Class members would also share knowledge of 'what can be done' to all these concepts or 'what they can do' when illuminated (viewed) from a single common perspective. This is what was meant by reference to the 'tool-box' above. Thus, whatever can be done to 'cars' can potentially be done to any 'Man-Made Objects' which makes it possible to talk about 'improving', 'manufacturing', designing', 'producing' and 'painting' or 'restoring' any such objects. By the same token, all "Commercial Objects' can be 'sold', 'priced', 'taken off the market', 'discounted', 'appraised', 'purchased' 'lose value' 'appreciate' 'cost' and so on and so on. Class of 'Commercial Objects" always makes its toolbox available to anyone who wants to communicate about cars from that particular vantage point. Just as a specific *aspect*, a *spectrum of aspects* and a *Class of Concepts*, this toolbox is also an element of the '*shared reality'*. It is now possible to look at its contents.

Inside the Toolbox

What we called a tool-box is a database of encyclopedic knowledge. It is indeed a dual phenomenon safely co-existing both in a cognitive and linguistic worlds for the benefit of any language user. This Cognitive-Linguistic Category (CLC) contains answers to the already mentioned three fundamental questions:

1) What can a concept referent *do* when viewed from a particular aspect?
2) What can be *done to* the concept referent when viewed from a given aspect?
3) What kinds or types of concept referents are there within a given aspect?

From a linguistic standpoint the tool box abbreviated as CLC is filled with three kinds of elements, three sets of word-forms that have the *property* of answering each of the questions above. In a more formal way, each linguistic element of the toolbox (CLC) must satisfy the criteria set by any of the three questions. In English, as in most Indo-European languages, these elements are 'noun-verb' (N+V), 'verb-noun' (V+N) and 'adjective/pronoun-noun' (Q+N) combinations. From a cognitive perspective, these are the true basic elements of *shared reality* and terminologically we gave them titles of *Operators* (Active and Passive) and *Qualifiers* (Gorbis, 2006 (a)).From any perspective we choose, these predicates are *units* of communication.

By itself, '*car'* is a meaningless word-form, a volume of unrealized potentialities. '*To drive a car*' contains or possesses 'meaning' because the unit extracts properties of 'Process' or 'Functional Object' from that volume. So does 'to own a car', 'to operate a car', 'to use a car' 'to subject a car to (what?) and other predicate units. 'Car' no longer remains a meaningless word-form but retains its collocation after-glow. Let us return to the earlier example: "*Living and working in California puts a terrible strain on your car"*. In a way the sentence is enigmatic. How did we get to decipher its core meaning

that "*driving a car on congested and ill-maintained roads in California subjects it to abnormal wear and tear"*. Those of us who live in LA know this without being reminded. But how does this text become meaningful to audiences unfamiliar with our roads and traffic? After all, the text contains no words like 'driving', 'operating', 'damaging'. 'damage', 'harm' 'roads', 'traffic', 'congestion' and so on. Within the theory of aspect semantics this cognitive deciphering process can be modeled in 5 steps as follows:

1) *"working and living*" primary Aspect – 'Living Object'; secondary Aspect – 'Operator'; tertiary Aspect – 'Human Operator'.
2) "*in California*" primary Aspect – 'Space'; secondary Aspect -- 'Territorial Location'; tertiary Aspect -- "Location Characteristics"
3) *"to put a strain on (what?)* (V+N) primary Aspect – 'Activity; secondary Aspect – 'Object Changing Activity'; tertiary Aspect – 'Harmful Results'.
4) "terrible strain" (Q+N) Primary Aspect – 'Effect Type'; secondary Aspect ''Abnormal'
5) "*your car'* (Q+N) primary Aspect – 'Material Object'; secondary Aspect – 'Moving Object'; tertiary Aspect 'Object of Possession'

Now the text and its real coded message can be written as follows:

Living Object	Space	Activity	Effect Type	Material Object
Operator	Territorial Location	Object Changing Activity	Abnormal	Moving Object
Human Operator	Location Characteristics	Harmful Results	n/a	Object of Possession

Here is the decoded message in a long-hand: *"Human Operator in Territorial Location that has some (unknown) characteristics (X) is in Activity which Changes Object of Operator's Possession beyond Harm normally expected to come to this Material Object."* And the likely solution for the riddle's variable X is '*Driving in California harms cars.'*

In the preceding pages we may have created an impression that in the course of cognitive-linguistic analysis an *Aspect* appears first and unit(s) follows like a toolbox carried by a workman to the job. This is not so and the opposite is most likely true. Each Operator and possibly each Qualifier must have a cognitive tag identifying it as belonging to or servicing a specific *cognitive Aspect*. For example, 'lousy work' summons the *aspect* 'Result'. 'work sucks' carries a tag of 'Process', but compare 'to ignore (whose?) work' which can service both 'Result' and 'Process' and thus requires textual inferences for disambiguation. 'To be late to work' is 'Time' *aspect* and so is 'to come to work (when?)' but 'to come to work drunk' identifies the *aspect* 'State or Condition'.

It is fair to admit that we do not know how real-life text processing works. We do not know if the grammatical structure analyses proceeds semantic or vice-versa. It may well

be that both proceed in parallel with some correlation of outcomes. (See Gorbis, 2005) However, it is very likely that any level of text processing is based on identification of universal grammar characteristics of the message and the metalingual '*aspects*' of concepts comprising the message or text.

What is a Cog?

"**COG©**" stands for "Cognitive Dictionary." Plainly speaking, it is a dictionary in which a concept is described or defined by its usage in a certain language or pair of languages.(See Gorbis, 1977). While many lexicographical works employ usage as a tool to refine a meaning or a shade of a meaning, efforts to create a systematic description of concepts solely by reference to their usages in a given language are unknown.

In the parlance of aspect semantics COG© is a database of '*concepts'* described by '*word-forms*' and their '*categories' (CLCs)*. 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 *Qualifiers*. 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 the *concept*.

Here are definitions of the basic terms and notions that were 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.

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 (CLC or 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.

This discussion should provoke (at least) two related questions that we are prepared to answer:

The first one is: where do we get '***aspects***' from? The second question is related: what comes first, the 'a*spect'* or the 'c*ategory'* which services it? (Linguistically, CLC c*ategory* is a database of *Operators* and *Qualifiers* servicing a given *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 *aspect* 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 *Qualifiers* 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 *Qualifiers*) for the selected *word-form* and then query each unit (*Operators* and *Qualifiers* 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* and an exhaustive list of cognitive -linguistic units servicing that aspect. That is the true objective of a Cognitive Dictionary.

Formal Rules of Concept Mapping for COG

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

Rule I. 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 (CLC) 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 Q+ N units that may apply to all concepts that share a given *aspect*.
Rule 5. By virtue of belonging to a CLC, a cognitive-linguistic category, each unit possesses a specific meaning with which it defines a property of each concept.
Rule 6. Each concept property and each aspect that invokes it is a universal metalanguage element.

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.)

COG approach allows us to introduce the notion of a *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.

References

Gorbis, Boris (1977) "Psycholinguistics and Generative Lexicography" in "Tetradi Perevodchika" issue 14, "Foreign Relations Publishing House" Moscow, 1977 pp103-116.

Gorbis, Boris (2005) "A Primitive Model of Metalanguage For Universal Grammar" Proceedings of MLMTA Conference, June 2005

Gorbis, Boris (2006 (a)) "The COG: How to Make Sentences From Concepts" in Proceedings of MLMTA Conference, June 2006

Gorbis, Boris (2006(b)) "Borrowing With Interest: Aspect Semantics View of Language Extension and Expansion" in Proceedings of MLMTA Conference, June 2006