Analyzability and Compositionality:
The effects of frequency of usage in Spanish proverbs

By: Manuel Burgos
University of New Mexico

“This incommunicable quality tells us
this sentence is proverbial and that one is not”
Archer Taylor, The proverb, 1931

This paper presents an analysis of Spanish proverbs, with the purpose of showing how frequency of usage has an impact on the configuration and evolution of these formulaic expressions. It proposes a definition of proverbs according to the semantic, syntactic and pragmatic characteristics of idioms (Fillmore et al, 1988). The examined data are proverbs of current use, extracted from Spanish virtual forums on the internet. The diachronic study of these idiomatic expressions will be based on two historical corpuses of Spanish and Refraneros (compilations of Spanish proverbs from 1620 until 1996). Using two of the more important concepts associated with formulaic expressions, compositionality and analyzability (Langacker 1987), their current forms and signification, as a consequence of their historical development, are described. This investigation uses proverbs as examples to demonstrate how frequency of usage affects grammar (Bybee, 2006, 2007).

I. Introduction

Functionalist Linguistics has brought a whole new set of perspectives to the science of language, most of which arise from a clear opposition to the traditional formalist assumptions. The point of departure of this paradigm is a conception of language understood as a phenomenological fact (Hopper, 1987) and not as an abstract and ideal system.

To place language in the dimension of interaction leads to considerations about the effects of usage in its configuration, and proposes the importance of empirical knowledge, or in the words of Bybee, “a usage based view takes grammar to be the cognitive organization of one’s experience with language” (Bybee, 2006: 711). Frequency is one of the most interesting phenomena that explain how usage affects
grammar. The repeated use of specific verbal units and constructions in natural conversations has important effects in several dimensions of language (Bybee, 2006, 2007; Bybee and Scheibman, 2007; Bybee and Brewer, 2007; Jurafsky et al., 2001; Erman and Warren, 2000; Hay, 2001).

One of the effects of the constant repetition of words is the tendency for them to be grouped together, which speakers will access as units. This will have consequences at a cognitive level, the way they are mentally processed by speakers (analyzability), and at a discourse level, where phonological, syntactic and semantic effects (compositionality) tend to occur. At the discursive level, one of the most evident tendencies is the use of formulaic expressions. These symbolic units have a gradient continuum, from schematic constructions (Croft and Cruse: 2004) to more fixed prefabs (Erman and Warren, 2000).

Within this continuum we find a special type of formulaic expression: idioms. These are considered “grammatical units larger than a word which are idiosyncratic in some respect” (Croft and Cruse, 2004: 230). This category contains subgroups according to their specific characteristics, which also place them in some point along the continuum. One of the subgroups are proverbs, which will be understood as a type of idiom with particular syntactic, semantic and discursive features (in the outline of the theoretical framework I will propose a characterization of them, according to these levels).

The study of proverbs has a long tradition in philology, phraseology, cultural studies, and obviously, the discipline that has them as object of study, paremiology. However, there has been little attention paid to these expressions from the perspective of functional linguistics. This paper attempts to be a contribution in this respect. Therefore,
it will take the analysis of Spanish proverbs in the light of a functionalist approach, specifically within the frame of the effects of usage in the emergence of grammar.

The objective of this analysis is Spanish Proverbs, *refranes*, focusing on their composition and diachronic change. Through the process of evolution of these units, and their current structure and meaning, it is possible to observe discursive and cognitive effects of frequency: compositionality and analyzability. Related to these two phenomena it is possible to observe some characteristics of what it is known as the ritualization of language (Haiman, 1994). At the compositional level we can see evidence of what Givon call *syntactization* in the forms of constructions (Croft and Cruse, 2004) and collocations (Koyke, 2001; Barlow, 2000), this last structure will be the result of frequency: the *chunking effect* (Bybee and Scheibman, 2007). In the analyzability level we will find evidence of *Habituation* and its linguistic manifestation of *semantic bleaching effect* (Bybee and Thompson, 2007) and *degrees of autonomy* in verbal units (Bybee and Brewer, 2007).

The order of presentation will be the following: first, I will propose a theoretical framework that places the study of idioms as evidence of the functionalist assumptions. In this same outline I will give a definition of *Refranes* as a specific class of idioms, according to the classification proposed by Fillmore et al (1988). Secondly, I will show the origin of my data and the methodology that I use to trace their linguistic evolution. Third, I will analyze their changes from the perspective of compositionality and analyzability. Last, I will conclude that proverbs and some of their syntactical, semantic, and pragmatic characteristics are the result of frequency of usage.
II. Theoretical Frame

Idioms are one of the phenomena of language that could not been explained satisfactorily by the generativist principles and therefore were placed in the lexicon dimension (Evaraert at al, 1995). These common expressions of language have been taken as a point of departure to criticize the formalist assumptions (Croft and Cruise, 2005) and sustain the functional approach, that basically propose that grammar is not the cause, but rather the consequence of language use (Hopper, 1987). Therefore, idioms have become an important research theme for functional linguistics (Erman, and Warren, 2000; Croft and Cruse; 2004, Fillmore, Kay and O’Connor, 1988; Evaraert at al, 1995).

The enormous literature on idioms has provided many insights about this verbal phenomenon, but at the same time, has created divergence in the assumption of its nature. That is the reason why we cannot find a general and satisfactory definition of the term, but different approaches that sometimes contradict one another (Makkai, Boatner and Gates, 1991). Taking this difficulty into account, I have decided to choose among the consulted authors the approaches and features I consider most relevant to this paper, without assuming this as a complete and absolute explanation of such a complex phenomenon.

Sometimes the most accurate definition is not the one that provides us with discrete features of the concept, but one that allows us to interpret several phenomena in accordance with a general framework. These definitions generally are broad and simple. I considerer that, in the case of idioms, it is better to start with one of this kind. Idioms will be understood in this paper as conventionalized complex expressions (Evaraert at al, 1995). These three words imply the most distinctive features of idioms. Convention, as
we will see, is probably the most important characteristic of these expressions. This means that in order for an expression to be considered an idiom, there must be a social agreement about its nature.

The next characteristic is *complexity*, which assumes two different levels, one in the structural dimension, where the syntax of idioms cannot always be explained under regular assumptions of syntactic formations (i.e. generativist rules). The second one is the semantic dimension, where the meaning or function of the idiom is not explained in terms of a direct relation between its components and the denotative meaning of each one of them, but rather a relation between the whole unit and a whole meaning or functions (Croft and Cruse; 2004; Barlow, 2000).

The third term, expression, refers to the communicative function of idioms. They work within the symbolic exchange of speakers, and are used to *express* not just in a descriptive way, but mainly to reveal some attitude of the speaker towards what is described. Their signification is dependent on a pragmatic context, and their use reveals what Jackobson proposes as the emotive or expressive function of language, where the election of certain phrases indicate what the speaker wants to tell about him/herself.

Croft and Cruse define idioms as “grammatical units larger than a word which are idiosyncratic in some respect” (2004: 230). And based on Fillmore’s analysis, they propose different types of idioms according to their characteristics in different levels. These levels correspond to:

1. The semantic interpretation of the idiom: where one type is *encoding*, which means that they can be interpretable in some respect by the meaning of the components (*answers the door*) or *decoding*, the ones that are completely
arbitrary, and not interpretable by their parts, but its meaning is based on the
convention among speakers (kick the bucket).

2. The “grammaticality” of idioms: where some of them will exhibit the regular
patterns of syntactical formation (Spill the beans, V + NP [DO]), and the ones that
behave according to a non-predictable syntactic rules, extragrammatical (e.g. … -
can’t passivize kick the bucket?).

3. The level of compositionality of idioms: where some are lexically filled, which
means that the components of the idiom do not admit any alteration in terms of
grammatical variation or the replacement of the units for others. These are known
as substantive idioms. The other type of idioms are those that admit grammatical
changes in their elements (i.e. change of tense or person in the verbs) and/or
interchangeability in the lexical components.

4. The pragmatics of idioms: this characteristic complements the first one regarding
the idiom’s semantic interpretation. In this case, the meaning of the idioms can be
related to its specific use in a communicative context, which will give it its
semantic force. The idioms that base part of their meaning within the context
where they are used are considered to have pragmatic point (Ex. Greetings. It is
different to use how are you in the first encounter with someone, than in the
middle of a conversation). The other class will be idioms that do not depend on a
particular context to be interpreted as an idiom (All of a sudden).

The fourth distinction about idioms allows me to introduce an approach to the
definition of the main concept of this paper: proverbs. The definition of this term has
been problematic in the consulted literature. There is not a general agreement about its
essential characteristics among the authors, and the majority of them initiate their papers or books with a comment about this difficulty, which demonstrates the level of uncertainty about its nature (Taylor, 1931; Colombi, 1989; Combet, 1993; Pérez Martínes, 1996)

To avoid here an extensive review of the different approaches to what are known as proverbs (Combet proposes in an humoristic way that there are as many definitions of proverbs, as proverbs themselves) or in the specific case of this paper (refranes: Spanish proverbs), I will consider refranes a special class of idioms, conventionalized complex expressions, and in accordance with the mentioned distinctions, the vast majority of them will be: First, encoding, in terms that their meaning will be not interpretable by the denotative relation of its parts, but it significance will be the result of the conventionalization of the phrase. On the other hand, their connotative meaning will be the result of a metaphor, which could have different degrees of interpretation, generally opaque (Feder Kittay, 1990). Second, extragrammatical, taking into account that the majority of proverbs are constituted by structures that do not correspond to regular syntactic rules. Third, substantive, proverbs are highly fixed compositions. They do not admit grammatical changes in their components, or the replacement of lexical items. However, as we will see in this paper, some of them will have variation from a diachronic perspective, which will demonstrate some level of openness, and also regional variation, where the same proverb its expressed differently regarding to a linguistic community.

And last, from a pragmatic point of view: proverbs express usually a socially accepted truth, which allow them to work as advice or to reinforce the veracity of an
argument (Pérez Martínes, 1996). This will make them dependent on a pragmatic context, where their interpretation is conditioned by the circumstances in which they are used.

III. Methodology and Data

Most of the studies about proverbs fail to mention the discursive contexts in which they were used. The absence of contextualization incurs in many problems. One of them is that we cannot be sure if the proverb it is actually used by speakers today, or if it is just a phrase that was used in the past. For example, many proverbs that appear in Refraneros (compilation of Spanish proverbs) are unfamiliar to speakers of contemporary Spanish. This difficulty creates a problem in terms of one of the basic characteristics of proverbs: conventionalization. This feature can be measured in terms of recognition by the members of a speech community, which means that conventionality is relative to specific linguistic environments (what is a proverb for me cannot be it for another native speaker of Spanish and vice versa).

To avoid this failure, the refranes that I analyze here are ones that I have noticed being used in current discourse. However, a single appearance of a proverb does not mean that it has a general use. Therefore, the selection that I did is based on the number of appearances in the two consulted sources, virtual Spanish forums.

There are two main sources for the proverbs that form the corpus of this study: one is Noticiero Digital (www.noticierodigital.com), which is the most active political forum in Venezuela, with a rate of more than 300 contributions per day (idem) and the other, El Tiempo (www.eltiempo.com), the main newspaper in Colombia, that also has an active forum, where readers express their opinions about the recent news. The discursive
function of proverbs as a mechanism to reinforce an argument makes these two spaces a rich environment for their appearance.

Over a period of one year, I extracted *refranes* from these sources, to compile a corpus consisting of 83 types. I have chosen the ones that have been used by most people, and I have based my analysis in the top ten according to the number of speakers that enunciate them in both web pages. Here I present a list of the ten most common *refranes* found in these pages, which are the examples on which I will base my analysis:

<table>
<thead>
<tr>
<th>Refrán</th>
<th>Number of appearances</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 No todo lo que brilla es oro</td>
<td>17</td>
</tr>
<tr>
<td>2 Del dicho al hecho hay mucho trecho</td>
<td>14</td>
</tr>
<tr>
<td>3 Camarón que se duerme, se lo lleva la corriente</td>
<td>12</td>
</tr>
<tr>
<td>4 A buen entendedor, pocas palabras bastan</td>
<td>12</td>
</tr>
<tr>
<td>5 Mas vale malo conocido que bueno por conocer</td>
<td>10</td>
</tr>
<tr>
<td>6 No hay mal que dure cien años ni cuerpo que lo resista</td>
<td>9</td>
</tr>
<tr>
<td>7 A caballo regalado no se le mira el colmillo</td>
<td>9</td>
</tr>
<tr>
<td>8 Lo que mal empieza, mal termina</td>
<td>7</td>
</tr>
<tr>
<td>9 El que se va para Barranquilla, pierde su silla</td>
<td>6</td>
</tr>
<tr>
<td>10 El que a hierro mata, a hierro muere</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1. Top ten *refranes* in the data

This ranking does not attempt to be evidence of the most popular *refranes* among native speakers of Spanish. They belong to specific discourse associated with the topics of the forums. For example, the forum of *Noticiero Digital* is a place dominated by the opposition of the Chavez government in Venezuela, therefore most of the *refranes* are use to sustain arguments that critique appearances (as in 1) false promises (as in 2) the finite existence of bad things (as in 6) etc. The same occurs in *El Tiempo*, where most of the *refranes* are used to make an argument about the political situation of Colombia. Our civil war is what most news generates, and the *refranes* are related with this topic, used for example to justify violent actions (as in 10), or to assume that a current situation is
better than a promised one (as in 5). This ranking is proof that refranes are relative to a discursive context (pragmatic point) and to specific communities of speech.

The second step of the study is to trace the historical development of these refranes. I used two different sources: 1) Historical corpora of Spanish and 2) Refraneros. The first source are corpora found on the web (www.corpusdelespanol.org) (CS) and (www.rae.com) (RAE), where there is evidence of its appearance in oral register (1900) and written registers (from 1600). The other source, Refraneros, are compilations of refranes, the oldest is from 1624 (Correas) and the most current is 1996 (Argos). I used nine different refraneros in total. I registered the different changes (if any) from each of the ten refranes in tables, with the mention of the source and the kind of change. Here it is one example of this diachronic change:

<table>
<thead>
<tr>
<th>Source</th>
<th>Year</th>
<th>Proverb</th>
<th>Kind of Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refranero de Correas</td>
<td>1630?</td>
<td>No hay mal que cien años dure, ni bien que a ellos ature.</td>
<td></td>
</tr>
<tr>
<td>Valera. Novel (CS)</td>
<td>1864</td>
<td>No hay mal que dure cien años</td>
<td>Syntactic: Change in the order NP + V x V + NP</td>
</tr>
<tr>
<td>9 Anonymous users of virtual forums</td>
<td>2007</td>
<td>No hay mal que dure cien años, ni cuerpo que lo resista</td>
<td>Syntactic, Lexical Semantic. Addition of a second part.</td>
</tr>
</tbody>
</table>

Table 2. Example of Change

The third step, and next section, is the analysis of these refranes and their changes under the principles of compositionality and analyzability. It has the main purpose of showing how frequency of usage affects the configuration of proverbs and their diachronic change.
IV. Analysis of Data.

Frequency has been shown to have important effects in the configuration of grammar (Bybee, 2006, 2007; Bybee and Sheibman, 2007; Bybee and Brewer, 2007 Bybee, Joan and Paul J. Hopper, 2001; Jurafsky et al., 2001; Erman and Warren, 2000; Hay, 2001). The studies based on frequency analysis tend to have a quantitative support of data for their hypothesis. This methodology has been made possible by new developments in technology, especially the access to enormous verbal corpora, tools of search and statistical analysis programs. These are, without question, an excellent way to prove the levels of frequency of specific tokens or types, but it does not mean that they are the only way to show frequency.

The study of proverbs can be an example of how frequent expressions cannot be shown in terms of quantitative analysis. First, proverbs commonly occur in oral discourse (Taylor, 1931; Combet, 1993). The oral data of the corpus is restricted to the appearance of the recording machine (195..?). Therefore the proverbs found in these corpuses are limited to oral conversation (their most natural environment) in the last 50 years; and the historical data of proverbs have their appearance in written texts, specially literature, which does not favor (with important exceptions as Don Quijote), these formulaic phrases.

The question then is: how, without quantitative support, it is assumed that proverbs have a high level of frequency? The answer to this question begins with an affirmation: frequency is a relative concept. It cannot be taken as a universal factor, but is related to specific linguistic context (speech communities, registers, conversations vs unilateral discourses, oral vs writing etc.). From this type of differentiation it is possible
to propose levels of frequency according to contrast. One example can be the phrase in Spanish *Sin embargo* (however) this phrase will have a low frequency appearance in the oral discourse in comparison to written texts.

Another consideration to take into account is that it is possible to think about two different kinds of frequency: synchronic and diachronic. The first analysis will measure the appearance of a given token or type in a specific period of time. The other will measure the appearance of the type or token contrasting several periods of time. This last kind of study has provided us with significant information about grammaticization and evolution of language (Bybee and Scheibman, 2007, Bybbe and Cacoullos, to appear). I based my assumption of high level of frequency of proverbs in the diachronic aspect from the fact that these proverbs can be found in refraneros published since the XVI century until the most recent ones (Argos, 1996). If the same proverb is registered for several authors that live in such distant periods of time, it is possible to infer that they have existed in speech for centuries, which will confirm its high frequency of use in relation with time.

Regarding it synchronic frequency, it has been proposed that “both written and spoken discourse are characterized by the high use of conventionalized sequences, which include sequences that we might call formulaic language and idioms” (Bybee, 2006: 713). One proof of this fact is related to the core characteristic of idioms: conventionality. In a previous work I did (Burgos, 2006) I tested several refranes among native speakers of Spanish at the University of New Mexico. I did a survey of recognition of these expressions with almost 60 speakers, and most of the refranes were familiar for more that 90 % of them. If those refranes were recognized for the vast majority, who came from
different countries, it implies that they were heard before in several linguistic environments, and therefore it is possible to conclude that they were used and heard enough times until they where stored by the speakers. This leads to the conclusion that proverbs have a relatively high level of frequency according to the familiarity of speakers with these expressions.

With these arguments I hope to have shown that proverbs have a high level of frequency in the diachronic and synchronic dimension. Now I will propose two notions of formulaic expressions associated with of frequency: compositionality and analyzability, that will be illustrated in the characteristics and evolution of some of the chosen proverbs.

V. Compositionality:
Langacker defines compositionality as “the degree of regularity in the assembly of a composite structure out of smaller components” (1987:457). The case of proverbs is a paradigmatic example of compositionality with a high degree of regularity. This regularity does not mean that they follow regular rules (i.e. generativist principles). On the contrary, and coherent with the extragrammatical characteristic of idioms, they tend to have structures that do not follow the regular patterns, as the next example illustrates:

<table>
<thead>
<tr>
<th>A caballo regalado</th>
<th>no se le mira</th>
<th>el colmillo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect Object</td>
<td>Verb</td>
<td>Direct Object</td>
</tr>
</tbody>
</table>

Table 3. Example of extragrammatical order

The regular pattern of a sentence in Spanish is: V + DO + IO. If the sentence would follow these regular rules, its structure will be: No se le mira el colmillo a caballo regalado (which would sound awkward to a native speaker). The regularity in this case is defined in terms of substantive idioms, which are those that do not admit grammatical modifications.
In terms of the formal proprieties of proverbs, it has been said that one of their main characteristics is their binary structure (Colombi, 1989). As we can see in the examples, each proverb contains two parts, subject and predicate (table 1). When two verbal units are repeated frequently there is a tendency to group them together. This process is known as chunking (Bybee and Scheibman, 2007). The authors propose that this constituency, degrees of relation between specific words used in the discourse, is in part a result of repetition.

The formal product of this process is generally known as collocations. These are understood as the frequent co-occurrence of lexical items (Koyke, 2001). This co-occurrence reflects one of the main characteristics of repetition: the most used combination of words tend to be stored as units, and this characteristic will make them easy to access in the speaker’s mind, allowing him/her to use them with more frequency, which will create a cycling effect (Bybee, 2006). Therefore, we see that collocations are a verbal product of language usage, and their level of constituency will be related to the level of repetition, or in Barlow’s words “speakers know about collocations, and collocations come in different strength in terms of how well routinised or entrenched they are in speakers’ linguistic systems”. (Barlow, 2000: 318)

In the evolution of proverbs it is possible to note how the two parts maintain a certain “solidarity” (Coseriu, 1981) in their changes. This means that when one of the parts changes, the other will tend to incorporate the changes to adopt the characteristics of the first one, as illustrated in the next example:
As we can see, the two components have a coincidence in the phonological expression of the last word. In the first registered form: *Fato – Rato*. In 1528, we see a replacement of words: *Fato* by *hecho*, and *rato* by *trecho*. It is plausible to think that the first change occurred in the first part, and that the second part adopted the change. However there is no evidence of this, as this would imply an intermediate point, with the change in the first part, while maintaining the older form in the second. In the consulted *refraneros*, it seems that the change was simultaneous. Interestingly, the changes occurred in the *refrán* tend to incorporate the sound [ch], this feature will be reinforced in the current example stage of the *refrán*, where *gran* was replaced by *mucho*, which leads this last form to have the sound in the four main words, which make it rhythmical and therefore more easy to remember.

The rhythm has been said to be a regular characteristic of proverbs (Taylor, 1931). However, just three of the ten chosen *refranes*, had this prosodic agreement:

<table>
<thead>
<tr>
<th>Source</th>
<th>Year</th>
<th>1 part of the Proverb</th>
<th>2 part of the proverb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sebastian de Orozco (CS)</td>
<td>1510</td>
<td>Del ditto al fato</td>
<td>ay grand rato</td>
</tr>
<tr>
<td>Feliciano de Silva (CS)</td>
<td>1522</td>
<td>Del dicho al fato</td>
<td>hay un gran rato</td>
</tr>
<tr>
<td>Don Quijote (CS)</td>
<td>1610</td>
<td>Del dicho al hecho</td>
<td>hay un gran trecho</td>
</tr>
<tr>
<td>Juan Varela (CREA)</td>
<td>1864</td>
<td>Del dicho al hecho</td>
<td>hay gran trecho</td>
</tr>
<tr>
<td>14 Anonymous users of virtual forums</td>
<td>2007</td>
<td>Del dicho al hecho</td>
<td>hay mucho trecho</td>
</tr>
</tbody>
</table>

Table 4. Example of “prosodic solidarity” between the parts.
Compositionality is not limited to formal features (phonological, morphological) but also involves semantics. Bybee and Scheibman propose “the more often two elements are used together, the more tightly they will be fused or bonded phonologically and semantically, and thus the tighter their constituency” (2007: 295). This principle applies to refranes, as the following examples demonstrate:

<table>
<thead>
<tr>
<th>No todo lo que brilla es oro</th>
</tr>
</thead>
<tbody>
<tr>
<td>A buen entendedor, pocas palabras bastan</td>
</tr>
<tr>
<td>Mas vale malo conocido que bueno por conocer</td>
</tr>
<tr>
<td>Lo que mal empieza, mal termina</td>
</tr>
<tr>
<td>El que a hierro mata, a hierro muere</td>
</tr>
<tr>
<td>No hay mal que dure cien anos ni cuerpo que lo resista</td>
</tr>
</tbody>
</table>

Table 6. Semantic Constituency

The components of these five proverbs tend to have a semantic co-relation. This could be seen as an example of the process of feedback in terms of production and cognitive recognition. On one hand it is possible to think that the words used have an internal meaning that relates them in terms of consequence, metonymy, antonomy, and synonymy. These semantic characteristic could be the explanation of their relation in the speakers mind, and therefore their constant repetition in the same utterance. On the other hand, those refranes that have this co-relation will be more easily memorized, as the result of blending of stored cognitive representations (Barlow, 2000).

In the case of the studied refranes, the cognitive representations can be expressed in the following description:

- **No todo lo que brilla es oro**: To shine is a property of gold.
- **A buen entendedor, pocas palabras bastan**: To understand is the result of hearing words
- **Mas vale malo conocido que bueno por conocer**: Bad as the opposite of good
- **Lo que mal empieza, mal termina**: To begin as the opposite of to end
- *El que a hierro mata, a hierro muere*: To kill as an active action opposite of being killed, the passive idea.
- *No hay mal que dure cien anos ni cuerpo que lo resistan*: The idea of something bad that lasts triggers the idea of the capacity of resistance.
- *El que se va para Barranquilla, pierde su silla*: To leave implies the idea of losing something.

One interesting factor is that of ten proverbs, nine of them have a compositional correlation: three in prosodic, and six in semantic. Just one of them have the two factors (*el que se va para Barranquilla, pierde su silla*). What can be concluded of this analysis of compositionally is that proverbs tend to have some kind of link between the parts (prosodic or semantically) that work as a connection that maintain the unity of the structure, and therefore will reinforce the ability of speakers to remember it, and consequently, to use it with frequency in discourse.

**VI. Analyzability:**

The cognitive dimension of the compositionality is analyzability. Langacker proposed that “the notion of analyzability is subtle. It does not refer to the intrinsic complexity of a structure, but rather to a person’s awareness of certain aspects of this complexity” (1987: 457).

One of the most important cognitive consequences of frequent repetition is that “word sequences are stored in memory and accessed as units” (Bybee, 2006: 714). This fact can be proven by some usage of idioms, where speakers provide just the first part of the expression, assuming that the hearer will complement it mentally and will access the whole meaning of the idiom. Barlow (2000) shows examples using the English idiom “it’s ain’t over ‘till the fat lady sings” and shows that there are many examples in which the part “*fat lady*” will be used as an element that triggers the notion of *to wait until the*
action ends to come up with conclusions. My data show many examples of proverbs where just one of the parts was used. The fact that the speaker (writer) may not have used the whole Refrán, implies that they assume the recognition of the whole by the hearer (reader) by one of its parts, which confirm the idea that they are accessed as units.

An important phenomenon related to this mentioned fact is that it is semantic bleaching, where “words or phrases that are much repeated lose their semantic force” (Bybee and Thompson, 2007). This is a constant theme in the analysis of idioms, where it is proposed that they are not understood by their parts, but by the whole form associated with a meaning or function (Croft and Cruse, 2005:257). In support of this argument, we find the experimental studies of Jennifer Hay, that show that speakers tend to better recognize the elements of a specific word if their components are more frequent than the whole. On the contrary, if the complex word is more frequent than their components, speakers will not recognize its internal structure and will access the word as a whole (2001).

Cruse addresses the idea that “words of the idioms are semantically empty” (quoted by Cacciary and Gluck 1995:44). Langacker proposes that “rather than seeing a composite structure as an edifice constructed out of smaller components, we can treat it as a coherent structure in its own right: component structures are not the building blocks out of which it is assembled, but functions instead to motivate various aspects of it” (Langacker, 1987:453).

In the analysis of my data I conclude that the interpretability of the proverb is based on two factors. The first one is coherent with the mentioned notions, and especially to what Bybee and Brewer (2007) call the autonomy effect. This means that the
construction loses its initial semantic force in terms of the literal associations, and
develops an independent meaning or function according to where and how the
construction is used. This is related to the notion that some idioms have a pragmatic
point. Proverbs have been studied according to their discursive function and the general
agreement is that they are commonly used to give advice, to recriminate some action, or
to reinforce an argument (Pérez Martínes, 1996; Colombi, 1989; Combet, 1993). This
means that they do not appear in isolation, but integrated into a general discourse, and
motivated by a pragmatic context.

I will exemplify this relationship between meaning and context with a memory of
a personal experience. The first time that I heard the refrán *A caballo regalado no se le
mira el colmillo* happened when I was a kid. My grandma visited us at our house, and
gave me a sweater as a present for my birthday. I complained about the color, and my
mom, in front of everyone, told me: *A caballo regalado no se le mira el colmillo*, with a
tone that was related to what I knew was recrimination. At the moment I did not realize
the internal meaning of the expression, but I understood clearly what she meant due to
her tone, nonverbal expression and the circumstances. The proverb got its signification
not by its internal structure, but by how and where it was used, which shows how it
developed a level of autonomy thanks to its link with a pragmatic point.

However, I do not consider that its meaning is completely arbitrary. Its internal
components reinforce or motivate the general idea of what the *refrán* is intended to
express. Some of them maintain their denotative meaning, and the other parts are
understood in figurative manner, that gives to the whole a metaphorical idea. The next
The table illustrates the semantic process, where the detonative will be considered *transparent* and the connotative *opaque*:

<table>
<thead>
<tr>
<th>COMPONENTS</th>
<th>A caballo</th>
<th>regalado</th>
<th>no se le mira</th>
<th>el comillo</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIGNIFICATION</td>
<td>opaque</td>
<td>transparent</td>
<td>transparent</td>
<td>opaque</td>
</tr>
<tr>
<td>INTERPRETATION</td>
<td>To any object</td>
<td>Given as a present</td>
<td>Do not look</td>
<td>Possible Defects</td>
</tr>
<tr>
<td>WHOLE IDEA</td>
<td>Do not look at the possible defects of any object given as a present</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Analizability Process

This analysis proposes that in order to interpret the metaphorical meaning of the expression an interaction between opaque and transparent elements of the composition is necessary. This idea is based on the assumption that there is certain degree of analyzability in the process of recognition and therefore usage.

Another argument that can be used to support the idea of certain analyzability is the diachronic and regional variations of *refranes*. One of the most common *refrán* found in my data could suggest this:

<table>
<thead>
<tr>
<th>Source</th>
<th>Year</th>
<th>Proverb</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enriquez del Castillo</td>
<td>1600?</td>
<td>El que se fue a Sevilla perdió su silla</td>
<td>Spain</td>
</tr>
<tr>
<td>Ruis Villamor</td>
<td>1998</td>
<td>El que se va a Sevilla pierde su silla</td>
<td>Spain</td>
</tr>
<tr>
<td>6 Speakers (T)</td>
<td>2007</td>
<td>El que se va para Barranquilla pierde su silla</td>
<td>Colombia</td>
</tr>
</tbody>
</table>

Table 8. Regional variation

If we assume that the lexical components of the idioms are “semantically empty”, and that these expressions have such a level of autonomy that the speakers do not recognize its parts, there would be no explanation for their change, due the fact that speakers would just reproduce the known phrase. In the example, we notice the variation of one of the lexical components according to the place (Spain/Colombia). The first and original *refrán*
uses the referent “Sevilla”, a well known city of Spain. In the Colombian variation, the city is replaced for one that is known in Colombia, Barranquilla, a name with similar phonological features. My hypothesis is that the replacement is a consequence of analyzability. Someone, in some point of time, decided to change a not well known reference (as is Sevilla for a Colombian) for a more familiar referent that played with the rhythm of the Refrán, to fulfill in a more effective way his/her communicative intention, which will be coherent with the idea of “grammar arises diachronically because of the commonly used discourse patterns that humans need to communicate” (Bybee, 2007:6). A counter argument could be that the change is the consequence of an accident: someone in some point was trying to use the original refrán, but he/she confused the location, and propose one that was more accessible to him/her, this could derive in expansion of this variation, and therefore the replacement of the original form.

As a conclusion of this section, it can be said that the context of use of proverbs is the main factor that determines their signification (they have a pragmatic point). The other factor is that this, plus their metaphorical meaning, has given them a level of autonomy in relation to their components. However, they are not absolute arbitrary, because their internal structure and meanings of the components reinforce the figurative interpretation. On the other hand, their regional and historical variation can be used as a proof of certain level of analyzability.

VII. Conclusions:

The study of formulaic expressions provides us with considerations about the effect of usage in the configuration of grammar. These expressions have a broad range of
forms according to a continuum that goes from open constructions (Croft and Cruse: 2004) to restricted prefabs (Erman and Warren, 2000).

Within this continuum there are idioms that are understood as *conventionalized complex expressions* (Evaraert at al, 1995), which are idiosyncratic in some respect (Croft and Cruse, 2004: 230).

Proverbs were defined as a class of idioms, with particular characteristics according to the typology proposed by Fillmore et al. (1988). In terms of their signification, they will be encoded, this means they cannot be understand by the denotative interpretation of their parts, but their meaning is based on the familiarity of the speaker with the expression as a whole. In terms of its grammatical features, they will have an extragrammatical configuration, which means that they do not always follow regular rules (i.e. Generativist principles) and they maintain a substantive structure; that is, they are usually fixed expressions that do not admit a change in configuration. Last, they will have a pragmatic point that is a link to a context that generates their meaning and/or function.

Based on proverbs used today, this paper shows how it is possible to interpret their form and meaning under the concepts of compositionality, “the degree of regularity in the assembly of a composite structure out of smaller components”, and analyzability, “a person’s awareness of certain aspects of this complexity” (Langacker 1987: 457). These concepts also were useful also to interpret their evolution, under the analysis of collocations, prosodic and semantic features, and their meaning in terms of cognitive effects such as semantic bleaching, where “words or phrases that are much repeated lose their semantic force” (Bybee and Thompson, 2007), and as a consequence of this there is
an autonomy effect (Bybee and Brewer, 2007). Regarding this last characteristic, it was proposed that even though the autonomy of proverbs in terms of their meaning and function is clear, there is a motivation in the components that reinforce the figurative meaning of the expression. This last assumption derived in the hypothesis that they have a certain level of analyzability, which could explain some of their temporal and regional variations.

When Archer Taylor, father of paremilogy, said “an incommunicable quality tells us this sentence is proverbial and that one is not” (1931), the quality he was probably referring to was familiarity, which is the individual recognition of conventionality, one of the results of frequency of usage.

**Bibliography**


Bybee, Joan and Cacoullos, Rena. To appear. The Role of prefabs in grammaticization: how the particular and the general interact in language change.


Corpus: [www.corpusdelespanol.org](http://www.corpusdelespanol.org) and [www.rae.org](http://www.rae.org)