Laura Pinnavaia, Nicholas Brownlees (eds.), *Insights into English and Germanic lexicology and lexicography: past and present perspectives*, 115-127 ©2010 Polimetrica International Scientific Publisher Monza/Italy

Lexical complexity: theoretical and empirical aspects

Gloria Cappelli

1. Introduction

In this essay we outline some theoretical and empirical aspects of the research carried out by the Pisa research group on lexical complexity. The notion of lexical complexity has been investigated from a specific angle – namely, starting from the assumption that languages are complex systems within which different types of structures act as organizers in order to make it possible for cognition to handle the immense amount of information involved in the communicative process. We have put forward the hypothesis that lexical items may themselves be viewed as complex dynamic microsystems which organize conceptual material in multiple ways depending on the task at stake. Within this view, which admittedly draws inspiration from the theories of complex dynamic systems elaborated by the empirical sciences, words act simultaneously as cues of mental representations, triggers of *ad hoc* conceptual constructions, and as anchors which prevent meanings from verging on the border of chaos. These hypotheses have proved substantial both in a translational and in a lexicographic perspective.

The essay is divided into four main sections: the first sketches an outline of the properties shared by theories of complex systems in different fields; the second provides the essential arguments supporting the view that the lexicon can indeed be conceived of as a complex dynamic system; the third presents a case study for lexical complexity and briefly discusses its implications for translation; the fourth presents the prototype of a lexicographic encyclopaedic file developed by the research group.

2. Complex dynamic systems

The term 'complex system' formally refers to a system consisting of many parts, at many different scales, which interact in a non-linear fashion (Bertuccelli Papi & Lenci, *2007*). Because they are non-linear, complex systems are more than the sum of their parts. A non-linear connection means that change on one side is not necessarily proportional to change on the other. In practical terms, this means that a small perturbation may cause a large effect (the so called 'butterfly effect'), a proportional effect, or even no effect at all. When there are many non-linearities in a system, its behaviour can be as unpredictable as it is interesting.

Besides non-linearity, complex dynamic systems exhibit the following crucial features:

a) they have a history: the history of a complex system may be important. Because complex systems are dynamic systems, they change over time, and prior states may have an influence on present states;
b) they may be nested: the components of a complex system may themselves be complex micro-systems;
c) their boundaries are difficult to determine: it can be difficult to determine the boundaries of a complex system. The decision is ultimately made by the observer;
d) their properties 'emerge'**:** emergence is the process of complex pattern formation from simpler rules;
e) they display adaptability: they adapt to external pressures;
f) they have internal organization.

Because of the complex interplay of all these properties, a complex system is one that by design or function or both is difficult to understand and verify. Before tackling the question of the lexicon and the reasons why it can be seen as a complex system, a few words on emergence are in order.

Emergence can be a dynamic process (i.e. a process occurring over time), such as the evolution of the human body over thousands of successive generations; or it can happen over disparate size scales, such as the interactions between a great number of neurons producing a human brain capable of thought – even though the

constituent neurons are not individually capable of thought. For a phenomenon to be termed emergent, it should generally be unpredictable from a lower level description.

Emergence is a central concept in the theories of complex systems, yet it is hard to define and very controversial. There is no scientific consensus about what weak and strong forms of emergence are, or about the extent to which they can be used to explain phenomena: the more complex the phenomenon, the more intricate are the underlying processes, and, therefore, the less effective the concept of 'emergence' alone is.

One reason why emergent behaviour is hard to predict is that the number of interactions between components of a system increases combinatorially with the number of components, thus potentially allowing for many new and subtle types of behaviour to emerge. For example, the possible interactions between groups of molecules grow enormously with the number of molecules, so much so that it is impossible for a computer to even count the number of arrangements for a system as small as 20 molecules.

The processes of emergency in the natural world seem to bear interesting resemblances to the process at work in meaning construal, so that this notion seems to be easily extendible to the processes operating in the lexicon.

3. The lexicon as a complex dynamic system

That complexity is an outstanding feature of the lexicon is proved by the many studies and approaches to word meaning. All of them testify to – albeit from often diverging theoretical assumptions – the intricate web of dimensions which need be taken into account when trying to explain how meaning is dynamically generated and comprehended in actual communication.

Here, however, we would like to push the idea of lexical complexity one step further, putting forward the hypothesis that complexity is not only the general category of thought most immediately evoked by studies in word meaning, but it also qualifies epistemologically as a defining property of the lexicon as a dynamic system in the sense defined above: the lexicon is characterized by dynamicity, non-linearity, nestedness, adaptability, self-organization

and stochasticity.[1] Within this perspective, we can view lexical complexity as essentially dependent on two main parameters acting as opposite pulling forces: the type and quantity of information required to describe the system, and the system's organizational properties.

The amount of information necessary to describe the system depends on the number of its possible states and on the regularity and predictability of its dynamics: the higher the randomness of the system, the higher its complexity (Bertuccelli Papi & Lenci 2007). This view of complexity applies both to the lexical system as a whole and to individual lexical items, which we hypothesize to be nested dynamic micro-systems differing for their degree of complexity and emerging with different properties from context-sensitive organizational pressures. Our claim is that words are pointers to conceptual structures (semantic spaces) out of which meanings are dynamically construed in context-sensitive modalities, following a non-linear process, and yet emerging in recurrent configurations with some degree of statistically relevant stability. This is the way the lexical macro-system of a language self-organizes in order to prevent communicative chaos.

Complexity in the lexicon can be evaluated along the two orthogonal axes suggested above, namely the type and number of dimensions dynamically entering into the meaning constitution process and the organizational properties. Concerning the former point, lexical meanings appear as inherently multidimensional entities. Meanings differ for the type and the number of their constitutive dimensions, as well as for their degree of interrelatedness. Moreover, the features defining a concept may come from very different domains – e.g. perception, motion, functionality, social reality, etc. – and may vary with respect to their salience. For instance, words like 'dog', 'cat', 'computer' and 'violinist' all refer to concrete entities. Nevertheless, in natural concepts such as those expressed by 'cat' and 'dog' perceptual features (e.g. colour, size, shape, etc.) are much more salient than in the meaning of 'computer' and 'violinist', which include a highly

[1] That the lexicon may be viewed as a dynamical system is supported by arguments from acquisition studies (lexical spurt), diachronic lexicology, and discourse lexicology; cf. Mac Whinney (1998).

prominent functional dimension. Similarly, the meanings of words such as 'school', 'theatre', 'book', etc. also add dimensions coming from abstract and social domains (e.g. information, time, etc.).

Word meaning multidimensionality is directly addressed by linguistic and psycholinguistic models of the lexicon. For instance, in the *Generative Lexicon* (Pustejovsky 1995, 2001) lexical entries are highly structured entities composed of multiple layers of information each pointing at different conceptual dimensions of meaning. Moreover, Vigliocco *et al*. (2004) bring empirical evidence supporting a model of the lexicon as a multidimensional space of integration of different types of features. The lexicon thus reveals a wide scale of complexity, which surfaces at the epiphenomenal level as polysemy, word context-sensitiveness, and so on. An overt correlate of lexical multidimensionality is actually provided by the different degrees of word selectivity in context (cf. Lenci 2005). Highly complex, multidimensional entities in fact determine multifarious word co-occurrence spaces, which in turn represent important probes to explore lexical multidimensionality.

Insofar as lexical meaning naturally involves multidimensionality, a theory of lexical complexity needs to spell out not only the nature of the dimensions which act as organizers of individual, contextual meanings, but also a set of higher order principles which may be hypothesized to determine the forms, dimensions, and status of the organization of meanings as represented in the mind.

4. 'Texture': A case study

The word we propose for a case-study is 'texture'. The meaning of this word is not easy to pin down, not so much because it changes with context – polysemy is to us an emergent property of the system – as rather because of the high level of underlying complexity it exhibits.

4.1 The lexical profile of 'texture'

A lexical profile (Stubbs 2001) for the word 'texture', based on a sample of 300 occurrences out of 900 in the BNC provides the following information in terms of co-selection, and semantic preferences:

TEXTURE

- Adjectives <good, soft, fine, light, firm, smooth, rough, coarse, crinkly, tactile, thick, rich, creamy, sticky, crunchy, glossy, thin, moisty, cool, meaty, greasy, spongy, velvetine, lovely, nice, fresh, poetic, discursive, dominant, traditional, polyphonic, symphonic, chromatic, musical, instrumental, open, bitable, heterogeneous, different, extra, uneven>;
- Nouns <sound T, surface T, madrigal T, colour T, skin > ;
- Noun of/in T <fullness of T, agglomeration of T, richness of /in T, delicacy of T, uniqueness of T, quality in T, transparency of T>;
- Texture of <butter, chocolate, harmonies and metres, political life, the words, some aspects of writing, the school, the skin, the story, a piece of old lino, cloth>;
- Verbs <improve T, smooth out T, feel the T, differ in T, T surges, cut out through T>;
- AND collocates <T and colour T and pattern, shape and T, T and flavour , taste and T , T and touch , odour and T , size and T , body and T, T and form, T and tone, T and shine, T and density, T and temperature, T and weight, T and feel >.

Analysis of the domains (literature, music, painting, cuisine, biology, computer graphics, linguistics) and co-texts in which the word occurs most frequently enables a finer identification of what 'texture' is and is not. Here follow some repeated co-texts:

- "they have a physical quality and that means *texture*";
- "a pianist has a fine ear for *texture*";
- "what is physical – colour, hardness, solidity, *texture*, smell, taste";
- "Watercolour papers differ in their absorbency, *texture*, weight and colour";
- "rock sequence, *texture* and composition".

The information we gather from corpus study therefore amounts to the following:

- 'texture' is a physical quality of objects;
- 'texture' denotes the properties held and sensations caused by the external surface of objects received through the sense of touch;

- 'texture' is sometimes used to describe the feel of non-tactile sensations;
- 'texture' can also be termed as a pattern that has been scaled down (especially in case of two dimensional non-tactile textures) where the individual elements that go on to make the pattern are not distinguishable;
- The actual meaning of 'texture' depends on the nature of the object considered. More specifically,
  a) in music, the word 'texture' is often used in a rather vague way in reference to the overall sound of a piece of music. A piece may be described as having a 'thick' texture, or a 'light' texture, or other terms taken from outside of music (e.g. Aaron Copland's more popular pieces are described as having an 'open' texture). The perceived texture of a piece can be affected by the number of parts playing at once, the timbre of the instruments playing these parts and the harmony and rhythms used, among other things;
  b) in cuisine, 'texture' is the feel of food on the tongue and against the teeth. Adjectives include 'crunchy', 'soft', 'sticky', 'mushy', 'hard', 'spongy', 'airy';
  c) in painting, 'texture' is the feel of the canvas based on the paint used and its method of application;
  d) in materials science, 'texture' is the property of a material's individual crystallites sharing some degree of orientation. It is seen in almost all engineered materials, and has a great influence on material properties;
  e) in soil science, soil 'texture' is used to describe the relative proportion of grain sizes of a soil or any unconsolidated material;
  f) in computer graphics, a 'texture' is a bitmap image used to apply a design onto the surface of a 3D computer model.

4.2 What this implies in terms of our view of lexical complexity

According to our definition, complexity is a function of the number and type of dimensions involved in the description of the (micro- or macro-) system, and is inversely related to the organization of a semantic space in terms of the stated principles, forms, dimensions

and states. From this point of view, which we call 'first order complexity', 'texture' turns out to be a complex word in English because of several factors.

First, its description makes use of multiple dimensions (e.g. visual perception, tactile perception, acoustic structure, surface appearance, solidity, etc.), each related to complex perceptual and crossmodal features. Second, 'texture' covers a multifaceted, fuzzy and loosely organized semantic space scattered through multiple domains (geology and material science, music, art and painting, food and cuisine, photography, computer graphics, etc.). More specifically, it would be difficult to identify a single frame to which the word may be referred. It might indeed be inserted in a 'perception' frame, albeit with a high number of idiosyncratic features; or, alternatively, it might be seen as a part of a 'physical quality' frame but again its placement within the frame could change with the nature of the object.

In terms of Merlini Barbaresi's principles (see Merlini Barbaresi 2003), 'texture' is obviously polysemous. Given its ambiguous orientation as a lexical pointer, it implies a low degree of indexicality. 'Texture' thus seems to be transparent as regards its morphological constitution (i.e. the root 'text' and the suffix '-ure'), but scarcely diagrammatic as to its conceptual matter: think for instance of the distance between such examples as 'the texture of Puccini's music', 'the texture of the skin' and 'the texture of painting'.

4.3 The 'translation problem'

Italian has no specific term to label the whole of the concept or the overall usage schema encoded by the English 'texture'. Italian, with a correspondence 'one-to-many' between the two linguistic systems, lexicalizes individual components of the schema and combinations thereof, dynamically foregrounding either

- perceptual modalities:
  a) *aspetto*: i.e. vision
  b) *grana*: i.e. touch (physical constitution + structure)
  c) *trama*: i.e. touch/vision (physical constitution + structure)
  d) *consistenza*: i.e. touch/ taste (physical constitution + structure)
  e) *tessitura*: i.e. vision/touch (physical constitution + structure)

or

- mental elaborations of the object perceptual properties:
  a) *struttura*: i.e. physical constitution and organization
  b) *essenza*: the core information accessible through the senses.

No individual Italian term is exclusively part of the semantic schema of 'texture'. Each option offered for the translation of 'texture'is polysemous, that is, it belongs to other schemata, and, if back-translated, might itself be a source of other complexities.

We believe that the problem cannot simply be dismissed by the observation of a lexical gap in Italian or of a one-to-many correspondence between one name for a concept and many names for different aspects or components of the concept. We take 'texture' to be a metaphor for the whole problem of the complexity of meaning, since it defines an interface between perception and cognition which the lexicons of English and Italian label not only in different manners, but with no comparable schematic regularity either. Hence the complexity of the lexical item 'texture' which, when observed through the magnifying glass of translation, turns out to be of two types. On the one hand, we observe '1st order complexity' in the mapping between words and concepts, and on the other hand, '2nd order complexity' is revealed in the cross-lingual mapping between word/concept pairs.

5. Prototype lexicographic file

The investigation of lexical complexity makes the limits of available lexicographic resources extremely evident. Part of the work carried out by the research group in Pisa consisted in creating a lexicographic file in the form of a website composed by hyper-textual pages that could make the complex interplay of the linguistic levels and of the contextual and cultural elements immediately visible. This file should ideally encompass the various dimensions that contribute to the complexity of a lexical item and show how networks of sense relationships are created, and what enters into the construction of meaning up to the level of implicit in a given stretch of text/discourse.

In order to build the prototype of the lexicographic file based on our theory of lexical complexity, we chose a lexical item pertaining

to the frame of vision, the verb 'see' (Bertuccelli Papi 2003). We endorsed the suggestions of Frame Semantics to define the dimensions relevant to the semantics of the verb, which were derived from the analysis of linguistic data retrieved from the *British National Corpus*. The relevant schemata for the frame of vision hypothesized are perception, cognition and affect which can be analysed in further lower-level schemata. The perception schema, for instance, would refer to conceptual dimensions such as agency, the temporal schema, the visual field, the object schema, the body schema, and the instrument schema. The cognition schema would involve conceptual information relative to attention, intentionality, awareness and purpose. The affect schema refers to such dimensions as quality, intensity and motivation. All these components are further subdivided into smaller 'components', thus, for instance, agency includes information about agent, causer, perceiver, etc.

As is evident, in the model of lexical complexity, the quality of dimensions involved is not sufficient to define the complexity of a lexical item, neither in terms of 1st order complexity nor of 2nd order complexity. Moreover, given that our purpose was to develop the prototype for a tool which, ideally, can assist translators too, the description of the lexical item was 'enriched' with further information of a different sort. Building on the assumption that an efficient translation requires the knowledge of a great variety of encyclopaedic and linguistic information, we tried to organize this information in such a way as to reduce the complexity of representation of lexical meaning. The file for the lexeme 'see' was built as to include:

- a phonological pointer
- a morphological pointer
- a syntactic pointer
- a semantic pointer
- a computational pointer
- a text/discourse pointer
- a historical pointer
- a varieties pointer
- translation tools.

Each pointer includes further subdivisions which show the organization of the vast amount of information needed to describe the semantics of a lexeme. Thus, for instance, the morphological pointer for 'see' includes examples of inflection, derivation (i.e. affixation, conversion and back formation) and compounding. The syntactic pointer lists the syntactic patterns in which 'see' can occur with examples from the BNC. It also includes information about marked structures – such as cases of ellipsis and null-object instantiations – and offers a link to the semantic pointer via the syntax-semantic interface. The computational pointer comprises information about collocations and collocational patterns as well as frequencies and salience of the collocates of 'see'. The variety pointer offers information about the lexical item both as a structural and semantic component of slang expressions. Other links provide dictionary entries (such as *Oxford English Dictionary*) and etymological information. Figure 1 shows the entry page for 'see'.

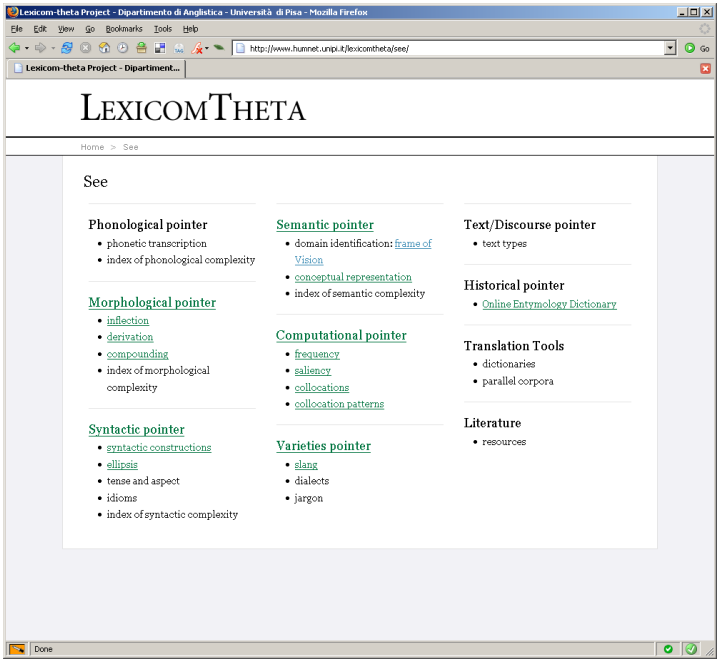

Figure 1.

6. Conclusions

As is evident, the compilation of such a file requires a great deal of work. Our aim was to create an operational framework founded on the theory of lexical complexity which can be developed by lexicographers into a precious reference tool both for linguists and translators.

Working at the file raised a number of questions, such as that relative to the possibility of quantifying lexical complexity. Ideally, each of the pointers should receive a complexity index based on the number and type of dimensions required to describe a specific domain. The complexity of a lexical item would turn out to be a function of these individual complexity indexes. The thorough investigation of these aspects and the possible applications of its results to the practice, didactics, and theory of translation are material for future research.

References

BNC *The British National Corpus* at http://www.natcorp.ox.ac.uk/.

Bertuccelli Papi, M. (2003) "Cognitive complexity and the lexicon", in L. Merlini (ed.) *Complexity in Language and Text*, Pisa, PLUS, pp.67-115.

Bertuccelli Papi, M. & Lenci, A. (2007) "Lexical complexity and the texture of meaning", in M. Bertuccelli Papi, G. Cappelli & S. Masi (eds.) *Lexical complexity: theoretical assessment and translational perspectives,* Pisa, PLUS, pp.15-33.

Lenci, A. (2005) "The lexicon and the boundaries of compositionality", in T. Aho & P. Ahti-Veikko (eds.) *Truth and Games: Essays in Honour of Gabriel Sandu*, Helsinki, Helsinki University Press, pp.119-138.

Mac Whinney, B. (1998) "Models of the emergence of language", *Annual Review of Psychology* 49, pp.199-227.

Merlini Barbaresi, L. (2003) "Towards a theory of text complexity", in L. Merlini Barbaresi (ed.) *Complexity in Language and Text*, Pisa, PLUS, pp.23-66.

Pustejovsky, J. (1995) *The Generative Lexicon*, Cambridge MA, MIT Press.

Pustejovsky, J. (2001) "Type construction and the logic of concepts", in P. Bouillon & F. Busa (eds.) *The Syntax of Word Meaning*, Cambridge, Cambridge University Press, pp.51-75.

Stubbs, M. (2001) *Words and Phrases. Corpus Studies of Lexical Semantics*, Oxford, Blackwell.

Vigliocco, G., et al. (2004) “Representing the meanings of object and action words: The featural and unitary semantic space hypothesis”, *Cognitive Psychology* 48, pp. 422-488.