

## On the duality of patterning<sup>\*</sup>

Jordi Fortuny

Centre de Lingüística Teòrica. Universitat Autònoma de Barcelona

Jordi.Fortuny@uab.cat

It is hard to discern any clear difference between “the search for language universals” and “the discovery of the place of human language in the universe”. They seem rather to be, respectively, a new-fangled and an old-fangled way of describing the same thing.

Charles F. Hockett (1963: 2)

### 1. Introduction

A well-known and uncontroversially universal property of natural languages is what Charles F. Hockett (1958, 1960, 1963) called duality or duality of patterning. This property can be expressed in the following terms: linguistic expressions can be decomposed into minimal meaningful discrete units commonly called morphemes that in turn can be decomposed into minimal distinctive though meaningless discrete units called phonemes. The duality of patterning is, thus, grounded on the discreteness of linguistic expressions and consists in the observation that minimal meaningful units are a combination of minimal distinctive units.<sup>1</sup>

If we consider, for instance, the English lexical morphemes *book* and *hook*, in their respective phonological transcription /bʊk/ and /hʊk/, it is easy to see that /b/ and /h/ are distinctive units, since the presence of one or the other determines a different lexical morpheme with a different meaning. Although /b/ and /h/ contribute to define morphemes with different meaning, they bear no semantic content; indeed an English speaker does not

---

<sup>\*</sup> I would like to thank Bernat Corominas-Murtra for his comments on a previous version of this squib. I also would like to express my gratitude to the editors for inviting me to participate in professor Jan Koster's Festschrift. This work was supported by the NWO-sponsored research Project *Dependency in Universal Grammar*. The final version of this squib is published in Mark de Vries & Jan-Wouter Zwart (eds.). *Structure preserved. Studies in syntax for Jan Koster*. LinguisticAktuell/Linguistics Today. Amsterdam/Philadelphia: John Benjamins Publishing Company.

<sup>1</sup> Hockett's term 'duality of patterning' corresponds to Martinet's (1960) 'double articulation'. We refer to Rosselló (2006) for an extensive presentation of Hockett's notion of duality of patterning and for a critical review of the usage of this term in the literature. Rosselló's article has inspired the main thesis of this work, namely "that the duality of patterning, as a lexicalization technique, merits to be considered as syntactic recursion in the sense that both mechanisms are on a par as manifestations of the 'infinite use of finite means' that characterizes language, to say it with Humboldt's words" (Rosselló 2006: 186).

have the ability to semantically decompose a lexical morpheme like /kæt/ ('cat') by assigning a meaning to each of its three phonemes.

The duality of patterning seems far from being an idiosyncrasy of the English language; it is to the best of our knowledge a property shared by all natural languages that have been described, regardless their typology or their areal or genetic classification. In Hockett's words, the duality of patterning is a plausible 'design feature' of language.

Hockett's (1960) general objective was to reopen the inquiry into the origin of language. In defining his program, he was admittedly guided by the comparative method; as he observed (p.5):

"Although the comparative method of linguistics, as has been known, throws no light on the origin of language, the investigation may be furthered by a comparative method modeled on that of the zoologist".

Hockett's guiding intuition was that one could set a short list of basic design features that were all present in all natural languages although no animal communication system displayed them all. After arriving at a particular list of thirteen features, he suggested that such a list could be used "to work out the sequence by which the ancestral system became language as the hominid—the man-apes and ancient men—became man" (p. 6).

In the remaining of this squib I will not be concerned with Hockett's lists nor will I attempt a review of Hockett's several works on the characterization and evolution of language, but I will focus my attention mainly on a single design feature, the duality of patterning, in order to study how certain conditions mainly related to mathematical information theory and structural complexity favor its emergence.

This brief work is organized as follows. Section 2 observes that the duality of patterning is a prerequisite for language to satisfy a basic principle of data compression. Section 3 argues that the duality of patterning or more generally the emergence of codewords is necessary to overcome transmission error in data transmission. Section 4 considers the possibility that a language neither displays the duality of patterning nor increases the number of distinct meaningful symbols, which leads us to consider how the duality of patterning and the unbounded syntactic procedure cooperate in order to reduce both the structural complexity of syntactic computations and the size of the morphological storage. Section 5 concludes this work by noting that an infinite number of linguistic expressions can be generated if the length of syntactic derivations lacks an upper bound and summarizing the main conclusions attained in the preceding sections.

## **2. Data compression**

We shall succinctly introduce the most basic elements of noiseless coding theory or data compression, a subcomponent of mathematical information theory (cf. Ash 1965, Cover and Thomas 2006), in order to show that a general type of duality is a prerequisite to define efficient coding algorithms.

Let  $\Omega = \{\omega_1, \dots, \omega_n\}$  be a set of messages to be coded by a coding algorithm  $C$ . As usual, we shall assume that the messages to be coded by  $C$  are generated by a discrete random variable  $X_\Omega$  ranging over  $\Omega$  with associated probability function  $p(w_i)$ , for  $(\forall w_i) (w_i \in \Omega)$ .  $C$ —as any algorithm—needs an alphabet, i.e., a finite set of basic symbols  $A = \{a_1, \dots, a_k\}$ .  $A^*$  will refer to the set of finite length strings of symbols of  $A$ ; for  $A = \{a_1, \dots, a_k\}$ , the four strings  $a_1, a_1a_2, a_2a_1, a_k \dots a_1$  would be elements of  $A^*$ .

Intuitively, a coding algorithm  $C$  proceeds by assigning to each element of  $\Omega$  an element of  $A^*$ . More precisely:

Let  $X_\Omega$  be a random variable ranging over  $\Omega$  and  $A^*$  a set of finite strings of  $A$ ; a *coding algorithm*  $C$  for  $X_\Omega$  is a mapping from  $\Omega$  to  $A^*$ .

A codeword is an element of  $A^*$  that is assigned by a coding algorithm to an element of  $\Omega$ ;  $C(\omega_3)$  is the codeword associated to  $\omega_3$ , and the set of codewords (a subset of  $A^*$ ) is called a code.

The objective of data compression is to code as much data as possible using the smallest number of digits. This is achieved by inversely correlating the length of the codewords with the frequency of the corresponding outcomes belonging to  $\Omega$  according to a given probability function, in such a way that the most frequent outcomes are assigned short codewords and the less frequent outcomes are assigned longer codewords. In sum, data compression attempts to minimize the expected length  $L(C)$  of a coding algorithm  $C$ , which is naturally defined as:

$$(1) \quad L(C) = \sum_{\omega_i \in \Omega} p(\omega_i) l(\omega_i),$$

where  $l(\omega_i)$  refers to the length of any codeword  $C(\omega_i)$ .

---

Table 1

$$\Omega = \{\clubsuit, \diamond, \heartsuit, \spadesuit\}$$

$$A = \{0, 1\}$$

$$A^* = \{x: x \text{ is a finite string of symbols of } A\} (|A^*| < \infty)$$

$X_\Omega$ with probability distribution $p$	$C_1$	$C_2$	$C_3$
$p(\clubsuit) = 2/4$	$C_1: \clubsuit \rightarrow 1$	$C_2: \clubsuit \rightarrow 1$	$C_3: \clubsuit \rightarrow 100$
$p(\diamond) = 1/4$	$C_1: \diamond \rightarrow 00$	$C_2: \diamond \rightarrow 101$	$C_3: \diamond \rightarrow 11$
$p(\heartsuit) = 1/8$	$C_1: \heartsuit \rightarrow 110$	$C_2: \heartsuit \rightarrow 0$	$C_3: \heartsuit \rightarrow 1$
$p(\spadesuit) = 1/8$	$C_1: \spadesuit \rightarrow 100$	$C_2: \spadesuit \rightarrow 00111$	$C_3: \spadesuit \rightarrow 0$

$$L(C_1) = (2/4)1 + (1/4)2 + (1/8)3 + (1/8)3 = 2/4 + 2/4 + 3/8 + 3/8 = 1,75 \text{ bits}$$

$$L(C_2) = (2/4)1 + (1/4)3 + (1/8)1 + (1/8)5 = 2/4 + 3/4 + 1/8 + 5/8 = 2 \text{ bits}$$

$$L(C_3) = (2/4)3 + (1/4)2 + (1/8)1 + (1/8)1 = 6/4 + 2/4 + 1/8 + 1/8 = 2,25 \text{ bits}$$

Table 1. Illustration of three coding algorithms  $C_1$ ,  $C_2$ ,  $C_3$ , each characterized as mapping from  $\Omega$  to  $A^*$ . The alphabet for the three algorithms is binary. ‘1’ and ‘101’ are two examples of codewords, the former associated to  $\clubsuit$  by  $C_1$  and the latter to  $\diamond$  by  $C_2$ .  $p$  assigns a probability to each element of  $\Omega$ .  $p$  is not uniform, provided that it does not assign the same value to each element of  $\Omega$ .  $C_1$ , but not  $C_2$  or  $C_3$ , minimizes its expected length by inversely correlating  $l(w_i)$  to  $p(w_i)$  for  $(\forall w_i) (w_i \in \Omega)$ .  $C_3$  directly correlates  $l(w_i)$  and  $p(w_i)$  and  $C_2$  does not correlate the two magnitudes whatsoever.

Clearly, the notion of coding algorithm is based on a general type of duality. Note the distinction between  $A$  and  $A^*$ ;  $C$  is a mapping from  $\Omega$  to  $A^*$ , but not from  $\Omega$  to  $A$ , and the elements of  $A$  are merely distinctive: they are combined to obtain the codewords that are assigned to the elements of  $\Omega$ , but they are not assigned to the elements of  $\Omega$  themselves. This is not an arbitrary property of the definition of coding algorithm: if there were no distinction between  $A$  and  $A^*$ , and the elements of  $\Omega$  were assigned directly to the elements of the basic alphabet, then all codewords would have the same length regardless the frequency of their respective outcomes; therefore, the average codeword length could not be minimized when the frequency of the elements of  $\Omega$  was not uniform.

What is relevant to our particular concerns is that the duality of patterning noted by Hockett seems a particular case of the general duality assumed in coding theory: the set of phonemes of a natural language corresponds to  $A$  and the set of morphemes approximately corresponds to the notion of code. By virtue of the duality of patterning, it is possible to assign different lengths to words depending on how often they are used, thereby minimizing the average length of words. It is a well-grounded observation, indeed, that “the length of a word tends to bear an inverse relationship to its relative frequency”, although such a

relationship may not be necessarily proportionate (Zipf 1936: 38). In this regard, it is also worth noting that the length of a word tends to diminish as its relative frequency increases along time, a tendency Zipf dubbed as the Law of Abbreviation.

As Hockett (1960) remarked, the Morse code also displays a duality of patterning, for it uses an alphabet of four symbols (a dot, a dash, a letter space, and a word space) in order to form finite length strings that are assigned to the letters of the English alphabet, in such a way that short sequences represent frequent letters and long sequences infrequent letters. Note that, in this case,  $\Omega$  is the English alphabet, whereas in the ordinary productions of a natural language  $L$ ,  $\Omega$  is a collection of basic semantic units (those that are assigned to the morphemes of  $L$ ).

Note, finally, that Saussure's (1916) arbitrariness is also a prerequisite for efficient coding. Let a sign be composed of a *signifiant* (English, 'signifier') and a *signifié* ('signified'). We say that a sign is arbitrary when the formal properties of the signifier do not mirror the conceptual properties of the signified, and non-arbitrary or motivated when they do. By extension, we shall say that a coding algorithm is arbitrary when the conceptual properties of the elements of  $\Omega$  are not mirrored by the formal properties of their respective codewords. A coding algorithm  $C$  that was non-arbitrary would lack the necessary flexibility to be efficient in the general case: it could not minimize  $L(C)$  in the case where  $p$  was non-uniform and the conceptual properties of some elements of  $\Omega$  determined that the length of their respective codewords was not inversely related to their frequency.

Arbitrariness is also a necessary condition to convey those meanings that cannot be connected in a non-arbitrary way to any strings, that is, those elements of  $\Omega$  whose conceptual properties have no correlate among the formal properties of the elements of  $A^*$ . Accordingly, non-arbitrariness may be a drawback for a coding algorithm to code certain elements of  $\Omega$  in the general case, provided that the set of properties of strings and the set of properties of the elements of  $\Omega$  may well be unconnected.

### 3. Data transmission

Let us consider the following reasoning, from Hockett (1960: 12):

"There is excellent reason to believe that duality of patterning was the last property to be developed, because one can find little if any reason why a communicative system should have this property unless it is highly complicated. If a vocal-auditory system comes to have a larger and larger number of distinct meaningful elements, those elements inevitably come to be more and more similar to one another in sound. There is a practical limit, for any species or any machine, to the number of distinct stimuli that can be discriminated, especially when the discriminations typically have to be made in noisy conditions".

It must be noted that Hockett's conclusion that the distinction between phonemes and morphemes appeared in order to avoid exceeding the practical limit to the number of distinct stimuli that any algorithm can discriminate, especially in noisy conditions, can be viewed as a consequence of Shannon's (1948) Fundamental Theorem for a Discrete Channel with Noise.

Shannon's concern was to increase the reliability of a communication system by designing codes that did not exceed the channel capacity, i.e., the maximum rate at which data can be transmitted through a channel with arbitrarily small equivocation. Shannon's Fundamental Theorem provides a sequence of codes of exponentially decreasing error probability with linearly increasing codeword lengths and exponentially increasing number of codewords. In this theorem, the existence of codewords is a prerequisite for the existence of codes that can overcome the transmission error as well as for increasing the number of objects that can be encoded.<sup>2</sup>

Accordingly, the emergence of codewords is motivated by elementary considerations related both to data compression and to data transmission. The emergence of codewords serves a double purpose: if we take into consideration data compression, the emergence of codewords allows to minimize the average codeword length, whereas if we take into consideration data transmission it allows to increase the length of words in order to avoid transmission errors.

There are at least a couple of qualms for the reasoning developed by Hockett in the above given passage that must be mentioned. Firstly it is not necessary to resort to the channel capacity (Hockett's practical limit) to understand why morphemes are distinguished from phonemes, provided that duality is a prerequisite for data compression, which is independent of how large the number of elementary symbols is or how close the number of stimuli is to the maximum number of stimuli that can be correctly discriminated. Therefore, the emergence in the evolution of language of the duality of patterning must have been favored as soon as language was a discrete system. Secondly, Hockett's reasoning does not consider the conceivable case where a language neither displays the duality of patterning nor increases the number of distinct meaningful elements up to exceeding "the practical limit (...) to the number of distinct stimuli that can be discriminated." We consider this possibility in the following section.

#### **4. The emergence of the duality of patterning and structural complexity**

It is hypothetically conceivable that a language displayed a system of around 40 sound units each of which corresponded to a particular semantic atom or monad; in such a language, a few sound units corresponding to monads would be combined to express non-atomic expressions that conveyed compounded meanings. This hypothetical language, which would not display the duality of patterning, would not differ from natural languages as to the degree of articulatory resolution and perceptual discriminability, although the structural complexity of the expressions of the former would be remarkably higher than those of the latter. More precisely, we may say that a language that did not display the duality of patterning and had the expressive power of natural languages would need to resort to the unbounded syntactic procedure responsible for combining minimal meaningful units to generate complex expressions much more extensively than natural languages. Such a syntactic procedure and

---

<sup>2</sup> Cf. Plotkin and Nowak (2000) for a reinterpretation of Shannon's theorem in their game-theoretic language model to account for the emergence of words in language.

the respective semantic rules that read the generated syntactic expressions are clearly at work in natural language, and in fact constitute central aspects in the modern formal study of grammar.<sup>3</sup>

The existence of the duality of patterning constrains the length or complexity of syntactic and semantic derivations by ensuring the existence of a set of minimal meaningful units that can be phonologically decomposed but not morphologically, in spite of conveying complex concepts, instead of monads. Although the duality of patterning reduces the complexity of grammatical compositions, note that resorting to the syntactic procedure is necessary to reduce the size of the morphological storage.

So far, we have indicated four advantages for the appearance of the duality of patterning:

- (i) it minimizes the number of distinct meaningful units, as noted by Hockett,
- (ii) it allows to overcome transmission errors,
- (iii) it restricts the resort to the unbounded syntactic procedure, and
- (iv) it minimizes the average length of words.

One could still add a further optimization condition that favors the appearance of the duality of patterning. There exist articulatory constraints for the combination of sound units into syllables; for instance, /*kpt*/ is not a possible syllable in any natural language, provided that any syllable needs to be headed by a [+sonorant] sound and neither *k*, nor *p* nor *t* is [+sonorant]. If *k*, *p* and *t* referred, respectively, to three different concepts *K*, *P*, and *T* that could be conceptually combined yielding a different concept [*KPT*], then such a new concept could not be expressed by the language under consideration. Therefore, constraints on what sequences of sounds can be articulated do not constrain the expressive power of a language if the elements on which articulatory constraints apply are meaningless.

## 5. Conclusion

In this squib we have been concerned with the observation that the distinctive units that ultimately compose linguistic expressions are meaningless, and we have provided arguments that show that this property has an important role in minimizing the complexity of linguistic expressions or computations as well as in overcoming transmission errors. In brief, the duality of patterning provides language with certain efficiency properties related to data compression, data transmission and structural complexity that ensure its feasibility.

We have assumed that the syntactic procedure responsible for generating hierarchically structured expressions whose terminals are morphemes lacks an upper bound, i.e., that in principle there is no constant or fixed natural number that defines the maximal possible length of an arbitrary linguistic expression. Accordingly, given an arbitrary expression whose length is the natural number *n*, the syntactic procedure can always construct a longer

---

<sup>3</sup> Cf. Fortuny and Corominas-Murtra (2009) for a set-theoretical definition of what we are here calling ‘the unbounded syntactic procedure’ and the basic structural relationships. This work capitalizes on Chomsky’s intuition that merge (internal and external) is the simplest way of generating hierarchically structured linguistic expressions (cf. Chomsky 2008).

expression whose length is the greater natural number  $m$ ; therefore the unboundedness of the syntactic procedure provides language with an infinite number of expressions that can refer to an infinite number of thoughts.<sup>4</sup>

This expressive power is not attained by directly combining a small set of meaningful sound units but a larger, though finite, set of morphemes. As argued, this allows to reduce the structural complexity of linguistic expressions. Resorting to the unbounded syntactic procedure is also necessary: a system with the same expressive power as language that resorted only to the duality of patterning would need an infinite storage of meaningful units.

This is, in sum, the picture we have arrived at of how the duality of patterning and the unbounded syntactic procedure work in tandem in order to provide language with a feasible design to make an infinite use of finite means.

## References

- Ash, Robert B. 1965. *Information theory*. New York: Interscience.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero and María Luisa Zubizarreta, eds., *Foundational issues in linguistic theory*, pp. 133-166. Cambridge: MIT Press.
- Cover, Thomas M. and Joy A. Thomas 1990. *Elements of Information Theory*. John Wiley & Sons, Inc.
- Fortuny, Jordi and Bernat Corominas-Murtra. 2009. Some formal considerations on the generation of hierarchical expressions. *Catalan Journal of Linguistics* 8, 99-111.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: MacMillan.
- Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203, 88-101.
- Hockett, Charles F. 1963. The problem of universals in language. In Joseph H. Greenberg, ed., *Universals of language*, p. 1-22. Cambridge, Mass: MIT Press. (Also in a revised version, in *The view from language: selected essays 1948-1974*, p. 163-186. Athens: The University of Georgia Press, 1977.)
- Martinet, André. 1960. *Éléments de linguistique générale*. Paris: Librairie Armand Colin.
- Plotkin, Joshua B. and Martin A. Nowak. 2000. Language evolution and information theory. *Journal of Theoretical Biology* 205, 147-159.
- Rosselló, Joana. 2006. Combinatorial properties at the roots of language: duality of patterning and recursion. In Joana Rosselló and Jesús Martín, eds., *The Biolinguistic Turn: issues on language and biology*, pp. 162-186. Barcelona: Promociones y Publicaciones Universitarias, S.A.
- Saussure, Ferdinand de. 1916. *Course de linguistique générale*. Paris: Payot.
- Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27, 379-423, 623-656. (Reprinted in Claude E. Shannon and Warren Weaver, 1949, *The mathematical theory of communication*, Illinois: University of Illinois Press.)
- Zipf, George K. 1936. *The psycho-biology of language*. London: Routledge.

---

<sup>4</sup> Note that, although there is no (fixed) upper bound, every expression may be of a non-identical and finite length. Thus, the cardinality of the set of linguistic expressions may be infinite although there is no infinitely long expression.