FACULTY OF ARTS
Charles University

A Code or Not a Code?
Metaphor, Metonymy and Metaphysics in Genetic Code Descriptions

Róbert Bohát, The Institute of the Czech National Corpus, Charles University; International School of Prague, rbohat@isp.cz

CZECH NATIONAL CORPUS

What?

Could the **cognitive metaphor theory** (CMT) help in clarifying the conundrum regarding metaphors and the genetic code? When it comes to **biosemiosis** (and the language used to describe it), *what is literal, what is metaphorical and to what degree?* From popular literature and textbooks to expert discourse, there is a wide mixture of (often inconsistent) use of metaphors and literal descriptions. Hence: **Is the genetic code really a** *code* **- or** *not* **a code? That is the question!**

Why?

Speaking of embryos and heredity in terms of "a book" goes as far back as ancient Psalms: "Your [God's] eyes looked upon my *embryo*, and everything was recorded *in your book*." (Psalm 139:16, ISV, emphasis added) Millennia later, Collins (among others) speaks of *The Human Genome Project* as "deciphering God's instruction book." (2006)
Controversy & Contradictions: The genetic *"code"* has been described as anything between as just a "metaphor carried too far" all the way to "literally true" and "more than a metaphor". (Weigmann 2004; Ridley 1999; Abel and Trevors 2006; etc.).
- "Although **the metaphors seem so intuitive** that it is hard for a geneticist to picture DNA as anything other than a code that transmits information, in her book *Who Wrote the Book of Life*, Lily E. Kay **doubted** that it is a natural **property of DNA** to be thought of as **informational** (Kay, 2000)." (Weigmann 2004, emphasis added)
- "The idea of the genome as a book is not, strictly speaking, even a metaphor. **It is literally true.**" (Ridley 1999; emphasis added)
Confusion: Many such confusing statements come either from biologists who do not seem to understand the profound complexity of conceptual metaphors, or humanities experts who seem to misunderstand the codon-to-amino-acid correspondence rules as just another 'chemical reaction' or some sort of crystallization. A **trans-disciplinary approach** might help.
So, was the optimism of hailing *The Human Genome Project* as 'learning the language of God' a statement of fact, a (mis)fitting metaphor or outright blasphemy? (cf. Weigmann 2004) Or could it be *something else*?

How?

Cognitive linguistics provides **at least two CMT-related methods** that can enable us to **systematize what is literal, what is metonymic and what is metaphorical** (and how) in biosemiotics.
The first method is the MIPVU (Metaphor Identification Procedure VU University Amsterdam), "based on extensive methodological and empirical corpus-linguistic research" with good inter-coder reliability. (Steen et al. 2010)
MIPVU (simplified) works as follows:
 "1. Read the text to get a general understanding of the meaning
 2. Determine the lexical units
 3a. Establish the **contextual** meaning of the unit
 3b. Determine if it has a more **basic** meaning [more concrete, body- related, more precise, historically older; not necessarily the most frequent meaning!] Does the contextual meaning contrast with the basic meaning but can it be understood in comparison with it?
 4. If yes, mark the unit as metaphorical." (MIPVU 2010)
 The *second approach* involves Dunn's (2015) methods of 'measuring' degrees of abstractness and metaphoricity; these could help in determining the basic meaning of "code", "information", "language", "alphabet", "letters" etc. from abstract and metaphorical contextual meanings with the *fact-status* and *function-status* decision trees, evaluating how much a given meaning depends on humans. Dunn's approach could supplement the MIPVU method.
The CMT would *inter alia* provide **a clearer distinction** between **metaphor** (cross-domain mapping of two unrelated domains) and **metonymy** (representation of one thing by a related [contiguous] thing). (cf. Littlemore 2015)

Ceci n'est pas un metaphore?

```
ORIGIN                                    ctcatgtggg
1 attaaaggtt                          cacgcagtat
  tataccttcc                         ttagtgcact
  caggtaacaa                  actc ggctgca
  accaaccaac               atctc ttgta
   ttctctaaa cgaactttaa aatc
   gctgtcactc ggctgcat
    tgcttacggt acacgt
      tatcagaggc
```

Ceci n'est pas un code?

Preliminary Findings

"The genetic code governs the exchange of messages on the cellular level." (Sebeok 2001)
Are these DNA "messages" *literal* **information? Yes.**
Basic meaning of "information": "In information theory …, a precise measure of the **information content** of a message, **measured in bits** … any data that can be **stored** in and **retrieved**…" (Collins Dictionary, emphasis added)
A specialized definition: "A message is a sign or **a string of signs transmitted** from a sign producer, or; source, to a sign receiver, or destination." (Sebeok 2001, emphasis added) In gene expression, **a string** of DNA triplet **codon signs** is **transmitted** via RNA codons to the string of amino-acids in a peptide. The genome *information content is measurable in bits* (Davies 2001: human genome = 6 billion bits = 750 Mb; would fill 5000 books.) Both DNA & language are mutually *transcodable* – from Venter's synthesized genome with quotes & team member names encoded through to DNA data storage & "genetic poetry" & Kac's "Genesis" (Callaway 2016; cf. Sýkora 2013; Kac 1999).
"Life's information-storage system [=DNA] is being adapted to handle massive amounts of information." (Lee 2019)
Is it a literal code? YES.
Basic meaning of "code": "A complicated system of rules, relationships, or instructions." (Macmillan Dictionary) Cf. in the genetic code, the DNA codon "TAT" **relates** to Tyrosine; TAT is a *sign* that *instructs* the positioning of Tyrosine in a polypeptide, the correspondence being a part of a **system of rules**. Cf. alphabetical "TET" & Morse Code "-.-".
A specialized definition: "A code is a set of rules that create a correspondence between two independent worlds." (Barbieri et al. 2012) "The genetic code is a correspondence between triplets of nucleotides, called codons, and amino acids. What is essential in all codes is that the coding rules are not dictated by the laws of physics. They are arbitrary in the sense that they are independent from physical necessity and this implies that they can be established only by natural or by cultural conventions." (Barbieri et al. 2012; cf. Barbieri 2008, Markoš et al. 2014)
cf. **Genetic Code**: "[T]he set of …codon-amino-acid correspondences, rules for the codon-amino-acid relationship." (Faltýnek 2011) So, if we accept the above definitions of "code", the genetic code **really is a literal code** (or perhaps *s-code*, Eco 1976).
So, is it a literal language? NO.
Both DNA & human language match the general as well as the technical definitions of a code. Does that imply that they are isomorphic? **It seems that** *not* **fully**. Hence, calling the genetic code "a language" would be a **metonymy**.
- **DNA/Language Similarities**: linearity, arbitrariness, dual articulation, (combinatorial) productivity… (cf. Faltýnek 2011)
- **DNA/Language Difference**: DNA "may not follow Zipf's law", typical for natural human languages (Tsonis et al. 1997; Matlach and Faltýnek 2016)
And is it a *literal* **text? YES (though DNA as a "book" would be a** *metonymy***).**
If a "text" is defined as Sebeok's 'string of signs transmitted from a source to a destination', then:
- Sýkora (2013): "[B]iological texts are not 'texts' only in a metaphorical sense, but in a full post-Lotmanian meaning of this term within literary studies. Genetic text written in DNA code is not a biological parallel to printed books, but rather a cybertext in the sense of Aarseth's definition of ergodic literature."
- vs. "The idea of the genome as a book is not, strictly speaking, even a metaphor. It is literally true." (Ridley 1999)
Similarly, these CMT approaches can be used to analyze the parallels (or lack thereof) between codons and letters, genes and words/sentences, genomes and books, protein function as the 'meaning of the code', etc. (cf. Matlach & Faltýnek 2016; Lacková et al. 2017))

So What?

Semiosis is "a necessary, if not sufficient, criterial attribute of life." (Sebeok 2001) Our semiosis about biosemiosis thus deserves to be taken seriously, without the misunderstandings caused by taking metaphors literally or literal facts metaphorically. Thus, having a **unified**, **systematic**, and **rigorous** set of methods (such as MIPVU and Dunn's) to distinguish the literal from the metaphorical (or metonymic) coherently and consistently can help decrease subjectivity and confusion. The **CMT can provide** biosemiotics with a useful and coherent **analytical framework** to help us deal with controversies, lack of conceptual clarity, ethical and metaphysical implications.
For example, once the basic meanings of both *language* and *genetic code* are established as types of **literal** *codes*, then there is a **contiguity** between the two. Hence, calling the genetic code "a language" would be a case of **metonymy** (or synecdoche). In other words, calling the set of codon-amino-acid correspondences a genetic *code* is a *literal* statement, but calling the genetic code "the **language** of the cell" would be **metonymic** (**not metaphorical**), as the two terms are non-identical but contiguous. Calling it "the language of God" would be a **metaphysical** extension of the **metonymy**, and not a "metaphor carried too far." Firstly, not all that is non-literal is metaphorical. Secondly, protesting or denying such theistic extensions on the grounds of philosophical **materialism** is as **metaphysical** as theism itself. The tension between the two schools of metaphysics should be dealt with metaphysically (and not by invoking science or metaphor beyond their 'jurisdiction'). After all, Polanyi (1968) argued: "As the arrangement of a printed page is extraneous to the chemistry of the printed page, so is the base sequence in a DNA molecule **extraneous to the chemical forces at work** in the DNA molecule." (emphasis added) 'Following the evidence wherever it leads' (Republic 394d) will sooner or later lead us *beyond* 'the chemical forces at work' and into the fields of history and metaphysics.
Prodi's Proposal: "Instead of thinking whether cells speak like us, the question should be asked whether we speak like cells." (Kull 2018) Since the appearance of the genetic code historically precedes (and is a *sine qua non* of) any human (or animal) codes, it seems that human codes are a (relatively simple) reflection of the much older genetic code. Indeed, **who mimics whom?**
Furthermore, could Wheeler's (1990) "it from bit" apply also to the role of the genetic code in the appearance of life: an "it" (cell) from "bit" (encoded genetic information)? Biosemiotics could become a robust **bridge** between **good science** and **good metaphysics**, helping us not only to avoid taking the literal metaphorically, but also to avoid mixing good science with poor metaphysics or invoking science where metaphysics should be used (and vice versa).
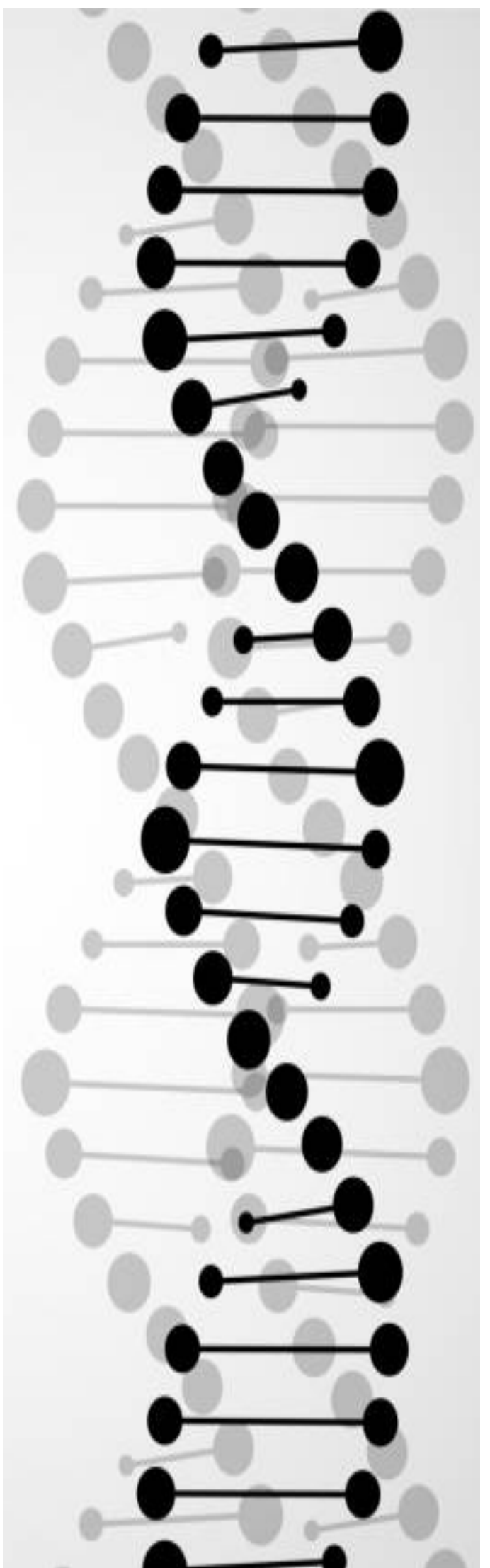
References

Abel, D., Trevors, J.T. (2006) "More than Metaphor: Genomes are Objective Sign Systems". *Journal of BioSemiotics* 2006 1(2): 253-267.
Barbieri, M. et al. (2012) "Code Biology: A Glossary of Terms and Concepts." *Code Biology Glossary*, International Society of Code Biology.
Callaway, E. (2016) "'Minimal' cell raises stakes in race to harness synthetic life". Nature 531, 557–558 (31 March 2016)
Collins, F.S. (2006) *The Language of God: A Scientist Presents Evidence for Belief*. Free Press, New York.
Dunn, J. (2015). "Modeling Abstractness and Metaphoricity". Metaphor and Symbol, 30:4, 259-289. Routledge Taylor and Francis Group.
Eco, U. (1976) *A Theory of Semiotics*. Indiana University Press, Bloomington.
Faltýnek, D. (2011) *Sémiotické primitivy v konstrukci gramatik*. Univerzita Palackého, Olomouc
Kac, E. (1999) "GENESIS". In: Gerfried Stocker and Christine Schopf (eds.), Ars Electronica 99 - Life Science . Vienna, New York: Springer, s. 310–313.
Kull, K. (2018) "Umberto Eco on the biosemiotics of Giorgio Prodi". Sign Systems Studies 46(2/3), 2018, 352–364.
Lacková et al. (2017) "Arbitrariness is not enough: towards a functional approach to the genetic code". Theory in Biosciences 136(1–2).
Lee, S.Y. (2019) "DNA Data Storage Is Closer Than You Think". Scientific American, July 1, 2019.
Littlemore 2015. *Metonymy: Hidden Shortcuts in Language, Thought and Communication*. Cambridge University Press, Cambridge.
Markoš et al. (2014) Biosémiotika II. Univerzita Palackého, Olomouc.
Matlach, V., Faltýnek, D. (2016). "Báze nejsou písmena". In *Studie z aplikované lingvistiky*, roč. 7, č. 1, s. 20-38. FF UK, Praha.
MIPVU (Metaphor Identification Procedure VU University Amsterdam). (2010). Metaphor Lab, VU University Amsterdam. Retrieved November 18, 2020.
Polanyi, M. (1968) "Life's Irreducible Structure". Science, New Series, Vol. 160, No. 3834, pp. 1308-1312, American Association for the Advancement of Science.
Ridley, M. (1999) *Genome*. Harper Collins, New York.
Sebeok, T.A. (2001) *Signs: An Introduction to Semiotics*. Second Edition. University of Toronto Press, Toronto.
Steen, G.J. et al. (2010). "A Method for Linguistic Metaphor Identification. From MIP to MIPVU". Amsterdam: John Benjamins.
Sýkora, P. (2013) "Od biologického kybertextu ku genetickej poézii". In: World Literature Studies 3•5 (22) • 2013 (39 – 53).
Tsonis, A. A., et al. (1997) "Is DNA a Language?" *Journal of Theoretical Biology*, vol. 184, no. 1, 1997, pp. 25–29.
Weigmann, K. (2004) "The code, the text and the language of God". EMBO Reports 5(2):116-8. EUROPEAN MOLECULAR BIOLOGY ORGANIZATION.
Wheeler, John A. (1990). "Information, physics, quantum: The search for links". In Zurek, Wojciech Hubert (ed.). Complexity, Entropy, and the Physics of Information. Redwood City, California: Addison-Wesley.