Acta Astronautica 116 (2015) 403–407

Contents lists available at ScienceDirect

Acta Astronautica

journal homepage: www.elsevier.com/locate/actaastro

Beyond an anthropomorphic template

John Elliott

School of Computing, Creative Technology & Engineering, Leeds Beckett University, Leeds LS6 3HE, England, UK

ARTICLE INFO

Article history:
Received 10 July 2014
Received in revised form
8 October 2014
Accepted 5 November 2014
Available online 13 November 2014

Keywords:
Language
Communication
Non-biological
Entropy
Ngrams
Lingdroids

ABSTRACT

In our endeavours to explore all possible forms that non-terrestrial communication may encompass, eventually we must throw off our anthropomorphic bias and investigate the implications of post-biological intelligence on SETI search strategies.

In the event a candidate signal is detected, our initial categorisation and assessment will focus on analysing its comprising constructs, to ascertain whether the information content is present; a fundamental signature of intelligence. To ensure our systems are capable of encompassing such intelligent communicators, we need to investigate both the contrasts and similarities of such non-biological communication and how this extends the known spectrum.

In this paper, we begin to investigate the likely signatures and contrasting structures such non-biological communicators may present to us, across a range of known machine communication phenomena, and discuss how such contrasting forms of information exchange can aid, extend and refine our detection and decipherment capabilities.

1. Introduction

In Shostak's paper 'What ET will look like and why should we care' [9], he highlights 'our anthropomorphic bias about extraterrestrials' and the implications of post-biological intelligence on SETI search strategies. Although the rationales behind searches to detect non-biological sentience are not our concern in this paper, our remit is to investigate the likely signatures and contrasting structures such non-biological communicators may present.

Above all else, we use communication [language] to convey information to someone or something else. Whether the conduit for this information is vocalised, written, or gestured, our purpose remains the same. In order that the message is understood, we use a shared 'code-book' of established [agreed] abstractions [symbols, sounds, and movements] to represent the meaning of our 'message', so the information is understood by the recipient.

Like any system, shortcuts can be made, when the resending of the message is 'cheap' and quick, or the context removes any possible ambiguity. Correction of incorrectly interpreted information or repetition of information, due to loss or 'damage' of comprising segments, can be costly; so, language evolves to negate (or at least reduce) such overheads, establishing rules and redundancies. Language is also structured according to the abilities [cognition; vocal dexterity] of the system users, to facilitate efficiency.

The natural communication system I describe is not a theoretic optimum but a highly efficient compromise. Humans typically cannot retain more than nine pieces of information at any one time, in their short term memory. Due to this limitation in our processing abilities, the communication system we have evolved uses inbuilt mechanisms [clauses, and phrases] to structure information in suitable chunks for processing. Nevertheless, the

E-mail address: J.Elliott@leedsbeckett.ac.uk

http://dx.doi.org/10.1016/j.actaastro.2014.11.005
0094-5765/

system is highly efficient and dynamic, having the capacity for infinity variety.

The 'forces' at work in evolving such a system of communication [language] is also shaped by the requirements of speaker and hearer [recipient]. In this, a form of reciprocal altruism is embodied, which follows a principle of least effort: compromising between the speaker's need to efficiently communicate the information and the hearer's need to receive and understand the message, unambiguously.

There is no common semantic assignment or common syntax: the veneer of communication. However, English is representative of a typical human language, in respect of its underlying structural signatures, for alphabetic [phonetic] based encoding, as well as representative of human language's entropic [information theoretic] signatures. It is the veneer of the sounds we utter and words we arbitrarily choose to assign semantic values and, the morphology system to 'glue' this together, that lends us to perceive the way we communicate differ significantly. When we strip this veneer away and look at the underlying structure, such as phrase chunking, where cognitive constraints operate, and internal structure of conditional probability, we then see the human language machine's [human brain's] common 'footprint' [5].

Our observations are admittedly from a single source: our own planet. However, extensive analysis of a wide variety of human variants of language has shown that they all adhere to the same underlying structures, dictated by the aforementioned constraints. It has also been possible to analyse the communication of a variety of other species, which arguably constitute 'aliens' on our own planet; for which results show the same 'forces' at work in their structure: results that support the principles of communication being 'universal' and therefore identifiable, where discovered. It is postulated, when communicating across the vast reaches of space written [text] communication is the most likely and is therefore the focus of analysis, for this paper.

2. Beyond the anthropomorphic mold

Nevertheless, in addition to the constructs of the natural language we use, with its evolved efficiency, range and flexibility, there are known alternative methods for representing information and knowledge. Given the possibility that such methods may be adopted, especially by non-biological forms, we need to look at the signatures such systems would present, to ascertain whether they are readily distinguishable.

One of the principle candidates, amongst these alternative methods for communicating information, which has been considered as viable for interstellar communication, is logic. Here mathematics meets semantics, in formal constructs, which convey information and semantic relationships in precise and explicit terms. Admittedly, I cannot recall any human actually using this method of communication, as a preference to natural language, but its potential for precisely [unambiguously] encoding and relaying information must make it a possible conduit for interstellar and remote intelligent inter-species communication. See Section 6 below.

Would a machine construct [evolve] a communication system based on logic or an optimised form of natural language encoding [no redundancy—100% pattern utilisation]? We will look at examples of known constructs in the machine [assembly] code lexicon and evolved robot communication from recent applied research, where the arbitrary pairing of a linguistic label to its assigned meaning is agreed by the robot community, without human intervention: The Lingodroids project (University of Queensland, Australia and Queensland University of Technology).

3. (Robot) Silicon chat

Lingodroids are robots, which use an onboard camera, sonar and a laser range-finder, to map the space around them [8]. This language, which sounds similar to the tones on a phone, is 'spoken' aloud by using a microphone and speaker. Experiments conducted in this project are a useful insight into how machine intelligence may develop communication.

The communication they use is not a typical computer [programming] language, but more of a human language. These words have been 'invented' by the robots themselves, using a variety of games to establish correlations between specific words and places, directions, and distances. And this includes teaching themselves brand new words for different lengths of time.

Although the Lingodroids described demonstrate the vital role played by communication for any task requiring more than one individual, the current state of evolution is nowhere near the complexity of a mature (fully developed) language, which can embed information (clauses, and phrases). Nevertheless, the basic concepts [building blocks] of language are developing, akin to those seen in animals, where relaying such information is vital for survival: where did you find the food? Where is the danger? Finding a mate, etc.

Examples of the Lingodroid's vocabulary show consistent use of short 4 letter words, to represent place names, distances and periods of time; significantly less word length variation than in human language. This supports the reasoning that machine language will develop their vocabulary, to explore all permutations within available variables, for maximum efficiency, as physiological and cognitive constraints will be negligible, in comparison to biologically based life forms.

> Geographical location examples: yifi, kiyi, gige, mira, xala, soqe, sihu, juhe, rije, pize, tuto, kopo, heto, qoze, yaro, zuce, xapo, zuya, fili.
> Distance measurement examples: puga, puru, vupe. duka, ropi, puga, huzu, hiza, kobu, bula.
> Temporal examples: kafi, puni, fohu, qija, fedi, tofe.

Unlike human discourse, where a given language only explores a subset of the phonetic space, typically, a given language will only use on average 50% of bigram (two letters) and less than 20% of trigram (three letters)