

# A MESSAGE FROM AFAR!

WHO ARE WE? ARE YOU THERE? WANT TO TALK?

## What's the Message?

### Here's your signal

The detection of a signal from another world would be a most remarkable moment in human history. However, if we detect such a signal, is it just a beacon from their technology, without any content, or does it contain information or even a message? Does it resemble sound, or is it like interstellar e-mail? Can we ever understand such a message?

This appears to be a tremendous challenge, given that we still have many scripts from our own antiquity that remain undeciphered, despite many serious attempts, over hundreds of years. – And we know far more about humanity than about extra-terrestrial intelligence...

We are facing all the complexities involved in understanding and glimpsing the intellect of the author, while the world's expectations demand immediacy of information. So, where do we begin?

### Structure and language

Information stands out from randomness, it is based on structures. The problem goal we face, after we detect a signal, is to first separate out those information-carrying signals from other phenomena, without being able to engage in a dialogue, and then to learn something about the structure of their content in the passing. This means that we need a suitable filter. We need a way of separating out the interesting stuff; we need a language detector.

While identifying the location of origin of a candidate signal can rule out human making, its content could involve a vast array of possible structures, some of which may be beyond our knowledge or imagination; however, for identifying intelligence that shares any pattern with our way of processing and transmitting information, the collective knowledge and examples here on Earth are a good starting point. For this reason, communication using all the types of human language, as well those used by some animals (e.g., dolphins, birds, and apes) has been studied, and even robots have received consideration.

Our natural language usually employs a hierarchy of structures and can be used in a spoken or written form, by means of characteristic sounds or sequence of symbols. For a language written by means of an alphabet, such symbols represent letters, and letters (or short sequences of letters) correspond to sounds. Letters combine to words, which carry lexical meaning, and words are put in relation by combining them to sentences following rules (of "syntax") in order to express statements. Alternative writing systems include using glyphs to represent whole words or syllables.

Latina	देवनागरी	ᱫᱷᱟᱱᱵᱟᱫ	한글
Ελληνικό	বাংলা	ไทย	ᄃᆞᆫ
Кириллица	తెలుగు	ລາວ	ᱥᱟᱱᱛᱟᱲ
תּיבּרַבּ	தமிழ்	ខ្មែរ	ᱵᱤᱨᱫᱟᱹᱜᱟᱲ
العربية	ગુજરાતી	ភាសា	ᱚᱛᱚᱴ
ᱠᱚᱰᱟᱨ	ಕನ್ನಡ	ᨆᩣ᩠ᩅᩣ᩠ᨾᩣ᩠ᨿ	漢字 / 汉字
ተፎጽኢዋ	മലയാളം	ᨈᩣ᩠ᩅᩣ᩠ᨾᩣ᩠ᨿ	ひらがな
ᱥᱚᱱᱚᱛ	ગુરમੁਖੀ	ᨆᩣ᩠ᩅᩣ᩠ᨾᩣ᩠ᨿ	カタカナ
Հայոց	ଓଡ଼ିଆ	ᨆᩣ᩠ᩅᩣ᩠ᨾᩣ᩠ᨿ	ᱠᱚᱰᱟᱨ
ქართული	සිංහල	ᨆᩣ᩠ᩅᩣ᩠ᨾᩣ᩠ᨿ	GWY

Contemporary writing systems: Latin, Greek, Cyrillic, Hebrew, Arabic, Syriac, Tifinagh, Ge'ez, Armenian, Georgian, Devanagari, Bengali, Telugu, Tamil, Gujarati, Kannada, Malayalam, Gurmukhi, Odia, Sinhala, Burmese, Thai, Lao, Khmer, Javanese, Sundanese, Batak, Lontara, Balinese, Thaana, Hangul, Tibetan, Modern Yi, Tai Le, New Tai Lue, (traditional/simplified) Chinese, Hiragana, Katakana, Inuktitut, Cherokee.

### Categorisation: The Universe of signals

All possible types of sounds and structures constitute a universe of signals. You can imagine it as a room, with random noise type structures at the outer edges, while at the centre one finds simple repetitive phenomena, such as clock ticks. In between these are areas of differing types of 'signal' structure complexity. In order that we can understand what type our unknown alien signal is (or might be) and where it fits within this universe of types, we need to model all the examples we know (both from Earth and from space) and where they fit in this signal universe. Ongoing research indicates that language has a special place. Although we may not know the words or what they mean, we can identify language from its structural signature. By understanding the signatures of all known information-carrying phenomena (e.g. images, language, mathematics etc.), we are given a baseline for initial categorisation.

### From signal to understanding content

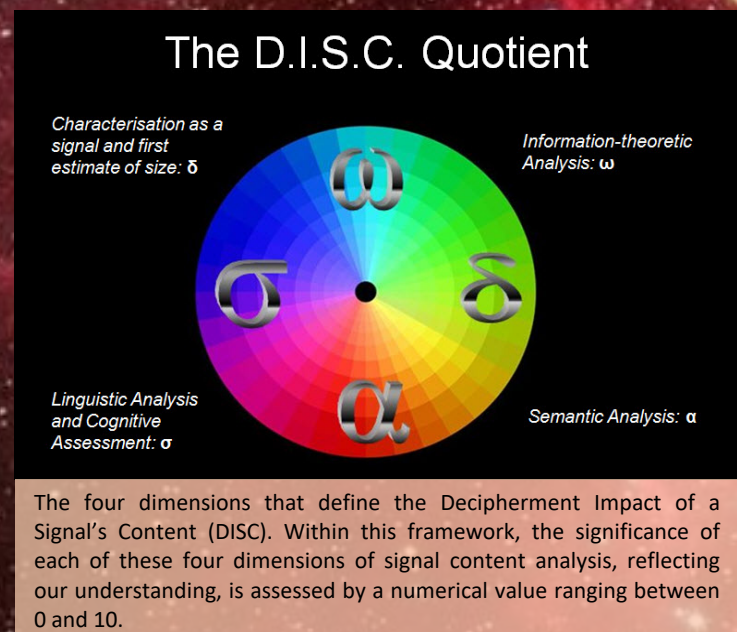
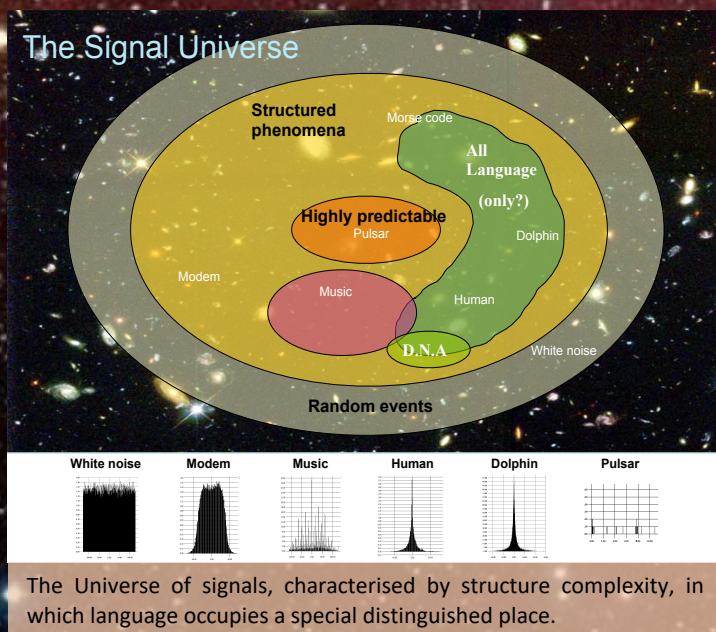
The Decipherment Impact of a Signal's Content (DISC) can be assessed along four dimensions, which represent key stages of analysis.

The more data we can capture, the more likely it will be that we can correctly interpret the content. The size of a message therefore is a first indication of its significance. Consequently, we would like to continuously gather and update our data received during analysis, if the signal persists or we subsequently capture additional transmissions from the same source. While a single tweet cannot exceed 280 characters (at most 1120 Bytes), all English Wikipedia pages constitute about 50 Gigabytes of data.

A first analysis of the signal's content would involve assessing its basic structure complexity, identifying its building blocks at the lowest and higher levels, and to measure their frequencies. Capturing how the abundance distribution of specific tokens is influenced by those of all other tokens is of particular importance.

Going beyond statistics, a further step would be a linguistic analysis that aims at identifying the functions of the identified building blocks. For example, in the English language, we distinguish different types of words such as nouns, verbs, and adjectives, and there are syntactic rules for combining them. Something to look out for in a message is a potential crib or primer. As we unveil information structures, we can then use known templates to help us look into the signal deeper but we can also use a method known as bootstrapping: a process that uses acquired information from features, to underpin and learn further knowledge of the phenomenon under investigation.

Finally, with structures being identified, we need to assign meaning to all elements. Such a semantic analysis involves looking at structural similarities within a document and probabilistic prioritisation of potential interpretations.



### The world is waiting

Besides the challenge of understanding the information contained in a signal from extra-terrestrial intelligence, we are also facing the question on what should be publicly disseminated. Should this be kept secret, and is that actually possible? Can the spread of news be controlled in the social media era? Or would full transparency be the best way forward? If we got a message, should we reply?

