Title: **Clinical markers from language complexity patterns**

Author(s): Barceló-Coblijn, Lluís[1]; Margarita Palmer-Riera[1,] Aritz Irurtzun[2 3]& Antoni Gomila[1]

Affiliation(s): UIB[1], IKER (UMR 5478)[2], CNRS[3]

Topic: Behavioural and Cognitive Neuroscience

Abstract: One of the first sources of information in psychological and psychiatric practice is language. Language can inform us about possible cognitive impairments, but also about atypical cognitive development. An example of the former is the logopenic progressive aphasia, attributed to either Alzheimer's disease or to frontotemporal lobar degeneration. An example of atypical cognitive development is the linguistic profile exhibited by people with an atypical genotype like Down syndrome or Williams syndrome. The way to approach the linguistic ability of patients with a potential linguistic impairment or atypical language development is usually based on tests (e.g., Boston test) and analyses of utterances by counting words (e.g., (in)capacity to perform a passive sentence or (in)capacity to say words that are semantically related like names of animals). However, as is well-known, sometimes the scores obtained are not fully informative, since some individuals reach scores at the edge of being classified as healthy, although the suspicion about a potential disturbing factor is evident from the speaker's spontaneous speech. There is the possibility to resort to other invasive techniques (e.g., lumbar puncture to detect abnormalities in the cerebrospinal fluid) and-or neuroimaging technologies (e.g. fMRI, DTI, EEG). We contend that there is a methodological gap between the basic tests and invasive/neuroimaging that can be filled by a data analysis technique that approaches speech data macroscopically (instead on focusing on single words/sentences).

We present a new technique for the extraction of more refined data from the language source using a combination of linguistic, morpho-syntatic analysis and network science. This methodology is able to extract complementary information about the speaker's linguistic system, that cannot be observed by the unaided eye. Complexity patterns of word combination emerge from a sample of spontaneous speech. These patterns take the form of a complex network made of words/morphemes (its nodes) and syntactic links (its edges), so that we can extract formal, objective indicators that speak about the cohesion of the speaker's linguistic system, which are the most connected words (hubs), which kind of links are crucial for the structure of her linguistic capacity. Therefore we consider that linguistic networks are good biomarkers – or endophenotypes – for language impairment or atypical language development.

We have applied this technique to 7 one-year-long corpora of typical children covering 6 different languages, 32 samples of Down syndrome, 20 cases of Specific language impairment and 20 cases of Hearing impariment. Moreover, we have the first data of a pilot study of Williams syndrome language network analysis. In sum, we observe that typical children develop their linguistic networks following a common developmental path, whereas the rest of atypical cases differ from that developmental path. The structure of atypical linguistic networks is different in many ways and the formal indicators support the idea that these speakers have a qualitatively different ability to combine words and information management.

The continued analysis by means of this technique provides a useful technique for tracking the (a)typical development of language or even its desintegration (in the cases of neurodegenerative diseases).