
Frequency, coding asymmetries, and the constant flow of linguistic information

Gertraud Fenk-Oczlon
Alpen-Adria-Universität Klagenfurt

Gertraud.fenk@aau.at

This presentation starts with some general remarks on the interrelationship between frequency and cognition: Relative frequency – overall token frequency as well as relative frequency within specific contexts – affects first of all cognitive processes. Higher frequency of use results in higher familiarity, faster accessibility or higher availability, and thus also in lower cognitive costs of speech production and perception. These cognitive mechanisms on their part influence linguistic variables such as length of morphological forms and even word order: More frequently units tend to be shorter (i) and to be placed before less frequently used units (ii).

I will present empirical results showing frequency-dependent asymmetric coding of morphological forms as well as frequency-dependent asymmetric ordering of linguistic units. Then I will argue – based on information theoretic considerations – that both rules (i) and (ii) contribute to a rather even distribution of information over the time, i.e. to a roughly constant flow of linguistic information:

(i) In terms of information theory, high relative frequency is related to low informational content. An element carrying a smaller amount of information can be processed within a shorter time. This means: Less time or less structural complexity for communicating less information. The proportionality function between information content and length of units provides an approximation to a constant flow of linguistic information.

(ii) As a sentence continues, the remaining words get more and more predictable – the number of possible and plausible continuations decreases, and so does the (subjective) information. To place informationally rich elements in a position, which is per se characterized by high information, would produce peaks of cognitive overload. An appropriate strategy to avoid such peaks is the tendency to begin a sentence or a clause with those words having a higher predictability in this context, e.g. with (groups of) words referring to (groups of) words of the preceding sentence, and with terms coding concepts activated by this preceding sentence. This tendency would explain, among other things, the

rule “old before new”, “topic before comment”, or “subject before object”.

AG5