Song lyrics and speech: similarities, differences and multi-dimension analysis of song lyrics form 1940 to 2009

Bértoli-Dutra, Patrícia (PUC-SP)

This presentation shows the results of a research aiming at finding the dimensions of linguistic variation present in Anglo-American popular music lyrics. The first goal was to find convergence of song lyrics speech and colloquial speech (general English) in order to highlight its relevance as a source for linguistic investigation. The theoretical framework of the whole research is founded by Corpus Linguistics, which is an area that is based on collecting and exploiting corpora, or a set of textual linguistic data carefully collected, in order to serve as a source for the study of a language or linguistic variety (BERBER SARDINHA, 2004, p.3). This first part of the study followed the principles of Corpus Linguistics (BERBER SARDINHA, 2004; BÉRTOLI DUTRA, 2002; HUNSTON & FRANCIS, 1999; SINCLAIR, 1991) first by describing the frequency of the words in the study corpus, then by describing the lexical-grammar patterns in the study corpus and finally by contrasting the patterns in the study corpus with lexical-grammar patterns present in general English. After collecting a study corpus of 1,078,882 words of song lyrics recorded by 30 different artists the word lists were extracted and contrasted with word lists (single words and trigrams) extracted from the BNC and the ANC. As a result we found that the most frequent single words in the study corpus are also relevantly frequent in the general English corpora. Besides that, only three out of the 500 most frequent trigrams in the study corpus do not occur in the other corpora, but they reflect something that we called "music language" (i.e. *"c'mon c'mon c'mon"; "oooh oooh oooh"; "oo oo oo"*). After finding the convergence of song lyrics and general English, we aimed at finding dimensions of variations of song lyrics according to Douglas Biber's model for a Multi-dimension analysis (1988), which presented a set of variation of American English. Biber's study assumes the probabilistic and functional characteristics of language (Halliday, 1991) and that linguistic variation occurs according to the context (BERBER SARDINHA, 2004; FIRTH, 1957; HALLIDAY; HASAN, 1989; HALLIDAY, 1991; HALLIDAY; SINCLAIR, 1991;WEBSTER; 2002). It also predicts that texts should be analyzed not only taking into account one but several linguistic features so as to determine their variation across linguistic functions. At the this part of the study, the corpus collected consisted of approximately 1,200,000 words from 6,290 song lyrics originally written in English. The corpus was tagged for its parts-of-speech features and for its semantic groupings. These features and the most frequent lexical bundles (3-grams) in the corpus and in general English (Google N-Gram corpus) were considered as variables for the factor extraction at the SPSS program. Factor analysis reduces the huge number of variables, grouping them according to their co-occurrence. This procedure is done through the identification of the distribution patterns of

variables. The 97 initial variables in our research were grouped into 13 grammar variables, 8 semantic variables, and 2 pattern variables (3-grams). Factor analysis resulted in three factors for each of those groups. The interpretation of the factors was conducted by focusing in three research questions: (1) what the main factors responsible for linguistic variation in song lyrics were; (2) which dimensions were expressed by the resulting factors; (3) how such dimensions were represented in relation to musical styles, to different artists and along the time. The factor extraction resulted in the existence of three factors. Grammatically they show the following oppositions (1) infinitive, gerund and modals versus nouns; (2) personal pronouns and possessives versus qualifiers; (3) verbs in the past versus verbs in the present. Semantically the factors show the predominance of (1) movement/time/speech/people/object; (2) markers of emotion and social acts; (3) markers of music manifestation. From the interpretation of the factors six different dimensions emerged the following dimensions: (a) argumentative versus informative; (b) interactive versus descriptive; (c) past narratives versus immediate context; (d) personal acts; (e) emotion and society; and (f) musical manifestation. The analysis of song lyrics on the dimensional scale showed how singers and bands, musical styles and the decade of the recordings are closer or more distant to each other in linguistic terms. The most representative style, artist and period of time for each of the dimensions, grammar and semantics, are as follows: (a) Punk Pop, Simple Paln, 2000's; (b) Rock'n'roll, Madonna, 1940's; (c) Country, Johnny Cash, 1970's; (d) Surf Rock, Beach Boys, 1960's; (e) Heavy Metal, Metallica, 1940's; and (f) pop Vocal, Franck Sinatra, 1940's. The Multi-dimension Analysis methodology cannot be considered as the only possibility for linguistic analysis of song lyrics or any other form of speech, but it served as a model for the research we present by making possible the contrast of linguistic features in functional terms. We were able to observe how songs are close or distant, similar or different according to their linguistic elements and not only according to their rhythm and musical style generally imposed by the media.