

## **POS ratios in the American presidential debates (1960-2008)**

### **A computer-based analysis of political speeches and parties**

The paper investigates the ratios of six parts of speech – adjectives, adverbs, determiners, nouns, pronouns, and verbs – in a specific kind of mediated political discourse, namely the American presidential debates. We claim that their distribution accounts for the diachronic evolution of public speeches and the influence of political party affiliation.

Our corpus is composed of twenty-two presidential debates organised between 1960 and 2008. These debates have been chosen because they reflect and evolve together with the society and the media practices. Also, they require the candidates to display very strong and easily recognisable identities. Finally, they provide an opportunity to study different strategies, depending on the attitudes candidates adopt towards facts and situations, and the specific relationships they build with their interlocutors.

Each debate has been transcribed; the data have been tagged with a morphosyntactic tagger (*TreeTagger*) and then automatically processed with a software of discourse analysis (*Hyperbase*). POS ratios are statistically determined (Z scores); their distribution is considered with multidimensional (or correspondence) analyses. Each quantitative description is illustrated with key sentences, *i.e.* excerpts considered as statistically representative of a candidate's speech. Two sets of analyses are carried out.

The diachronic study shows a grammatical evolution: nouns, determiners and adjectives (statistically) prevail from the 1960s to the 1980s, while verbs and pronouns are dominant during the following two decades – adverbs do not appear as a relevant item here. Interestingly, this is also true for French literary data (see the works of E. Brunet), *i.e.*, another language family, morphosyntactic typology and genre of discourse. Therefore, POS ratios don't result from English linguistic features; we suggest that extralinguistic variables may be relevant here; particularly, for our corpus, the evolution of social and media practices. In a word, pre-1980s audiences call for conceptual and descriptive speeches whereas post-1980s audiences favour dynamic and interactive speeches (see U. Eco's analyses) which put the stress on the actions and the personal commitment of politicians. The 1976 and 2004 elections are quite representative of this dichotomy.

Candidates also embody their political affiliation and adopt a different strategy according to their (challenger or incumbent) position (see the works of R. Friedenberg). Therefore, their grammatical profiles may not be coherent with the diachronic study. *E.g.*, R. Nixon favours the use of verbs and A. Gore the use of nouns: hence, they don't meet the 1960s and 2000s linguistic expectations. Here, we suggest to consider the political affiliation. Indeed, our analysis links POS ratios to the attitudes Republicans and Democrats want to exhibit: while the former seem to prefer subjective commitments made explicit with mental and modal verbs, the latter tend to build a factual speech, thus displaying a sense of objectivity towards situations.

In sum, the paper studies POS frequencies with a quantitative method used in corpus linguistics and discourse analysis. Our investigation of the American presidential debates shows that grammatical choices may be influenced by extralinguistic variables: the social and media practices, the political parties' attitudes.

[488 words]