PART TWO

ARABIC VARIATION AND THE MEDIA
ARABIC ORAL MEDIA AND CORPUS LINGUISTICS: 
A FIRST METHODOLOGICAL OUTLINE

Marc Van-Mol
Katholieke Universiteit Leuven

So far, the term Media Arabic has not clearly been defined. Often scholars refer to media Arabic as a specific kind of Arabic, without a precise delimitation of its boundaries. In some cases, newspaper Arabic is understood by the term (Versteegh 2003: 138 & Holes 1995 who often refers to contemporary journalistic MSA). Arabic media language, however, comprises at least also its spoken counterpart that is used in radio and television programs. Research on the ‘oral’ segment of media Arabic, so far, has rather been limited to news broadcasts (Harrell 1960; Van Mol 2003). Media Arabic, however, covers a much larger field than the two above-mentioned segments. In order to investigate media Arabic on satellite television we have recorded in total 2100 hours of Arabic satellite television in the year 2000. A great part of the content of these satellite programs has been transcribed and tagged.

The compilation of this ‘oral Arabic language corpus’ is part of a larger project that is called MARC-2000 which is the abbreviation of Modern Arabic Representative Corpus 2000, a synchronic corpus which has been compiled at the Katholieke Universiteit Leuven in Belgium. The aim of the compilation of the MARC—2000 corpus is to obtain an objective scientific view about the Arabic language situation and its characteristics such as it was used in the year 2000 in its most varied settings.

In this article we will give an intermediate report about our empirical findings related to the methodological problems of classifying media language in general and oral media language more specifically. Our provisional inventory of oral media Arabic gives an idea of the complexity of oral Arabic on satellite television. We present some methodological issues we face in trying to define precise classifications for oral media Arabic language use in order to ensure an optimal scientific research on the sampled material.

As has clearly been stated by Blanc (1960) and later in more detail by Badawi (1973) spoken Arabic forms a continuum of different layers.