Active lexical knowledge plays an important role in the evaluation of linguistic competence. In L2, active lexical knowledge is often investigated in terms of the lexical richness of written or oral texts. However, this global term lexical richness covers two principal dimensions: lexical diversity (the relation between number of types and tokens) and lexical sophistication (the use of difficult words). Since the 1930’s researchers have been developing several algorithms measuring the first dimension (lexical diversity) with the main goal of avoiding negative text length influence. As a side effect, the definition of what these algorithms actually measure has sometimes been neglected. More recently, scholars (e.g. Daller & Treffers-Daller, 2003; Malvern & Richards, 2007; Jarvis, 2013) have tended to assess lexical richness by taking into account different features of the words (e.g. frequency).

In this talk, we will argue for a combination of measures of lexical diversity with measures of lexical sophistication to evaluate the active lexical knowledge of bilingual children. We will discuss existing measures as well as their applicability for short texts written by children. This discussion will be based on the results of a pilot-study in which we described the lexical diversity of 105 short texts by Portuguese migrant children in Switzerland (aged 8 to 10) measured with algorithmic measures (HD-D and MTLD) and assessments by untrained raters. The results showed that quantitative measures and subjective ratings of lexical diversity are correlated, but that the untrained raters were also influenced by other characteristics of the text (e.g length, or ”quality”). In the second part of the discussion, we will present an instrument in development, adapted from already existing tools, to assess the lexical richness of children short written productions in French, German and Portuguese combining diversity and sophistication.

Keywords: lexical sophistication, migrant children, lexical diversity.