

Difficoltà con la lemmatizzazione dei binomi lessicali – come potrebbe cavarsi d’impaccio la fraseografia moderna?

Damir Mišetić

Résumé : Le tournant pragmatique des années soixante-dix, ainsi que l’intérêt croissant pour la langue parlée au cours des dernières décennies, imposent à la phraséographie et à la lexicographie modernes de développer des approches nouvelles et mieux adaptées à la lemmatisation des expressions figées, omniprésentes non seulement dans la langue littéraire et l’écrit en général, mais surtout dans la conversation quotidienne. Une classe d’expressions figées – à savoir les combinaisons phraséologiques – particulièrement intéressante non seulement pour la recherche phraséologique est représentée par les binômes lexicaux, présents dans toutes les langues modernes, y compris dans les langues romanes, et parmi elles l’italien, bien que de manière moins fréquente que dans les langues germaniques et slaves. Leur lemmatisation pose néanmoins d’importantes difficultés aux lexicographes et phraséographes, car elle exige à la fois une solide formation théorique et des compétences pratiques en matière d’identification et de traitement, en particulier dans le cas des binômes réversibles. Dans ce contexte, la linguistique quantitative pourrait s’avérer d’une grande utilité, comme le montrent des recherches récentes menées sur plusieurs langues européennes. La phraséographie et la lexicographie italiennes, toutefois, semblent rester à l’écart de ces développements. S’il est vrai que l’empirie, à elle seule, ne peut résoudre tous les problèmes qui freinent le développement de la recherche phraséographique moderne, il n’en demeure pas moins qu’elle peut nous fournir des données fiables sur le patrimoine phraséologique des langues européennes en général, et de l’italien en particulier.

Mots-clés : binômes lexicaux, corpus web, lemmatisation, irréversibilité, italien