

KIParla: una risorsa modulare per lo studio dell'italiano parlato

Caterina Mauri (Università di Bologna)

L'intervento si propone di presentare KIParla come infrastruttura di ricerca dedicata alla raccolta, organizzazione e documentazione di parlato spontaneo in italiano contemporaneo, illustrandone l'architettura, i criteri di costruzione e le prospettive di riuso.

Attualmente il corpus comprende oltre 350 ore di registrazione e circa 3.200.000 token trascritti, distribuiti in moduli differenziati per contesto comunicativo e profilo sociolinguistico dei partecipanti. KIParla è progettato secondo un modello pivot-based: ogni modulo è costruito secondo criteri metodologici omogenei, così da garantire comparabilità interna e integrazione progressiva di nuovi dataset. L'unità minima di organizzazione è l'evento comunicativo, descritto attraverso metadati strutturati che includono parametri sociolinguistici dei partecipanti e caratteristiche situazionali dell'interazione. Le trascrizioni, realizzate in ambiente ELAN, registrano fenomeni tipici del parlato (sovrapposizioni, interruzioni, segmentazione turnale) e sono concepite per consentire livelli successivi di annotazione linguistica.

Saranno illustrati i criteri di selezione dei contesti, le scelte di bilanciamento tra varietà e controllabilità dei dati e le modalità di accesso e riuso della risorsa. KIParla si configura come corpus cumulativo, estensibile e metodologicamente trasparente, finalizzato a sostenere analisi empiriche sull'italiano parlato contemporaneo, in prospettiva sia qualitativa sia quantitativa.