

# Pentingnya kecakapan numeris, sains data dan komputasi bagi peneliti bahasa abad 21

Gede Primahadi Wijaya Rajeg

Program Studi Sastra Inggris, Fakultas Ilmu Budaya, Universitas Udayana

<https://orcid.org/0000-0002-2047-8621>

## Abstrak

Salah satu sub-target yang tercakup dalam tujuan utama ke-4 SDGs (“Pendidikan Berkualitas”) adalah peningkatan literasi dan kecakapan numeris (sub-target 4.6) (United Nations 2022). Tujuan utama makalah ini adalah mengulas pentingnya kecakapan numeris, khususnya statistik, serta kecakapan terkait seperti sains data (*data science*) (Donoho 2017) dan komputasi (mis. pemrograman) (Bryan & Wickham 2017), bagi penelitian ilmiah dalam bidang bahasa. Stereotip yang lumrah di kalangan peneliti bahasa terkait dualisme kualitatif-kuantitatif/statistik adalah bahwa bahasa merupakan fenomena kualitatif sehingga tidak terlalu memerlukan pendekatan kuantitatif dalam pengolahan dan analisis data (Gries 2009: 4). Hal ini berimplikasi pada tidak dicantumkannya mata kuliah (MK) ilmu statistik dasar dan sains data dalam kurikulum pembelajaran bahasa di berbagai jenjang (setidaknya di lingkungan Fakultas Ilmu Budaya [FIB], Universitas Udayana, namun tidak untuk Prodi S1 Antropologi di FIB yang memiliki MK “Pengantar Statistik”, dan Prodi S3 Linguistik yang menyisipkan statistik dasar pada MK “Linguistik Korpus”). Ketimpangan metodologis tersebut kini mulai berkurang mengingat kajian bahasa abad 21 telah memasuki revolusi kuantitatif (Janda 2017) seiring dengan perkembangan teknologi dan keberadaan data kebahasaan digital melimpah (*big language data*). Lebih lanjut, ilmu pengetahuan (*science*) secara umum, dan bahasa khususnya, bergantung pada kesistematian dalam pengumpulan data empiris guna menguji teori dan/atau mengajukan teori/hipotesis terkait bahasa. Statistik dan sains data berperan penting dalam penelitian ilmiah karena keduanya adalah ilmu tentang pengumpulan dan memaknai data (kuantitatif) (Donoho 2017). Data (kuantitatif) yang dikumpulkan bisa berasal dari hasil angket ataupun data kalimat melimpah yang keseluruhannya dianalisis secara kualitatif kemudian disarikan secara kuantitatif (mis. dengan frekuensi atau proporsi dan persentase), untuk peneliti kemudian ajukan generalisasi dalam kaitannya dengan teori/hipotesis terkait bahasa. Makalah ini akan mengulas dua contoh sederhana: (i) pentingnya evaluasi statistik terhadap data kebahasaan sebelum mengajukan generalisasi terhadap suatu fenomena kebahasaan, seperti pemakaian umpatan oleh pria dan wanita, fenomena sintaksis, dan semantik leksikal; dan (ii) pengelolaan data kebahasaan pada MS Excel.

**Kata Kunci:** Statistik, Sains Data, Komputasi, Linguistik Korpus, Penelitian Bahasa