

Untuk data distribusional berikut ini, hitung nilai Chi-Square, nilai efek (Cramér’s V/Phi), dan tentukan apakah distribusi tersebut signifikan secara statistik.

1. Gries dan Stefanowitsch (2004) menemukan bahwa kata kerja *give* muncul 461 kali dalam konstruksi Dwitransitif (mis. *Billy gave Diane a present*) dan 146 kali dalam konstruksi Datif preposisional dengan *to* (mis. *Billy gave a present to Diane*). Konstruksi Dwitransitif muncul dengan kata kerja lain sebanyak 574 kali sedangkan Konstruksi Datif dengan *to* muncul dengan kata kerja lain sebanyak 1773 kali.
 - a. Hitung frekuensi harapan kata kerja *give* di kedua konstruksi tadi.
 - b. Tentukan apakah distribusi tersebut signifikan secara statistik.
 - c. Seberapa besar nilai efek distribusi bentuk dan makna tersebut?
 - d. Gunakan spreadsheet Excel **2x2** untuk distribusi ini.
2. Rajeg, Rajeg, & Arka (2020) menemukan bahwa (i) kata *mengenakan* digunakan dalam arti ‘memakai (mis. pakaian)’ sebanyak 1011 kali, dalam arti ‘melaksanakan (mis. pajak, denda)’ sebanyak 48 kali, dan dalam arti ‘lainnya’ sebanyak 28 kali; (ii) kata *dikenakan* digunakan dalam arti ‘memakai (mis. pakaian)’ sebanyak 171 kali, ‘melaksanakan (mis. pajak, denda)’ sebanyak 253 kali, dan dalam arti ‘lainnya’ sebanyak 18 kali.
 - a. Adakah perbedaan signifikan dari distribusi makna dengan bentuk morfologis kata dengan akar kata *kena* tersebut?
 - b. Makna manakah yang berasosiasi (i.e. muncul lebih sering dari yang diharapkan secara) signifikan dengan salah satu bentuk morfologis kata berakar *kena* tersebut?
 - c. Seberapa besar nilai efek distribusi bentuk dan makna tersebut?
 - d. Gunakan spreadsheet Excel **2x3** untuk distribusi ini.
3. Rajeg, Denistia, & Rajeg (2018) melihat asosiasi antara bentuk morfologis predikat dan dua konstruksi negasi di Bahasa Indonesia yaitu, *tak* dan *tidak*. Asosiasi ini diukur dengan menghitung jumlah tipe predikat berskema *di-X-kan*, *me-X-kan*, dan *ter-X-kan* yang dinegasi oleh *tidak* dan *tak* dalam konteks konstruksi [*tak/tidak* + predikat]. Ditemukan bahwa *tak* menegasi 29 tipe predikat *me-X-kan*, 21 tipe predikat *di-X-kan*, dan 20 tipe predikat *ter-X-kan*. Kemudian, *tidak* menegasi 327 tipe predikat *me-X-kan*, 246 tipe predikat *di-X-kan*, dan 21 tipe predikat *ter-X-kan*.
 - a. Kombinasi negasi-dan-bentuk morfologis manakah yang muncul lebih sering dan jarang dari yang diharapkan?
 - b. Apakah distribusi tersebut signifikan secara statistik?
 - c. Seberapa besar nilai efek distribusi tersebut?
 - d. Rancangan Excel sheet manakah yang digunakan untuk distribusi di atas?
4. Konstruksi [*will/I’ll/shall*+INFINITIVE] (mis. *I will/I’ll/I shall learn statistics*) muncul 3184 kali pada korpus LOB (British English) dan 2945 kali pada korpus BROWN (American English). Konstruksi [*going to*+INFINITIVE] (mis. *I am going to learn statistics*) muncul 172 pada korpus LOB dan 144 pada korpus BROWN (Stefanowitsch 2004).
 - a. Apakah penutur Bahasa Inggris Amerika menggunakan konstruksi [*going to*+INFINITIVE] lebih sering (dari yang diharapkan) secara signifikan dibandingkan dengan penutur Bahasa Inggris British?

Gede Primahadi W. Rajeg (Prodi Sastra Inggris, Fakultas Ilmu Budaya, Universitas Udayana, Bali) (2020, 3rd November)
Materi pendukung untuk sesi “Statistik untuk Linguistik Korpus” yang dibawakan pada lokakarya *Kebahasaan dan Kesastraan untuk Kelompok Kepakaran Layanan Profesional (KKLP) Perkamusan dan Peristilahan, Sekretariat Badan Pengembangan dan Pembinaan Bahasa* (4 – 5 November 2020)

5. Kombinasi ajektiva-nomina *little girl(s)* ditemukan 3100 kali; *little boy(s)* ditemukan 2000 kali; *small girl(s)* ditemukan 100 kali dan *small boy(s)* ditemukan 440 kali (Stefanowitsch 2004).
 - a. Hitung frekuensi harapan untuk masing-masing kombinasi ajektiva-nomina.
 - b. Apakah distribusi **2x2** ini signifikan secara statistik?
 - c. Bayangkan kita menemukan kombinasi *little child(ren)* sebanyak 1000 kali, lalu *small child(ren)* sebanyak 2250 kali. Ulangi kalkulasi di bagian (a) dengan tabel **2x3** di Excel sheet.

Referensi

- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on “alternations.” *International Journal of Corpus Linguistics* 9(1). 97–129.
- Rajeg, Gede Primahadi Wijaya, Karlina Denistia & I Made Rajeg. 2018. Working with a linguistic corpus using R: An introductory note with Indonesian negating construction. *Linguistik Indonesia* 36(1). 1–36. <https://doi.org/10.26499/li.v36i1.71>.
- Rajeg, Gede Primahadi Wijaya, I Made Rajeg & I Wayan Arka. 2020. Corpus-based approach meets LFG: Puzzling voice alternation in Indonesian. *figshare* (Paper Presented at the 25th International Lexical-Functional Grammar). [figshare](https://doi.org/10.6084/M9.FIGSHARE.12423788.V3). <https://doi.org/10.6084/M9.FIGSHARE.12423788.V3> (15 October, 2020).
- Stefanowitsch, Anatol. 2004. Quantitative thinking for corpus linguists. Tutorial. http://www-user.uni-bremen.de/~anatol/qnt/qnt_dist.html (4 August, 2011).