

Chi-Square Test – Goodness-of-fit

Gede Primahadi Wijaya Rajeg

MK Linguistik Korpus

Prodi Doktor Linguistik, Fakultas Ilmu Budaya, Univ. Udayana
Oktober 2021

Chi-Square Test – Goodness-of-fit

- Untuk variabel kategorikal/kualitatif (e.g., nominal dan biner)

Chi-Square Test – Goodness-of-fit

- Menguji apakah **frekuensi pengamatan** untuk kategori dari suatu variabel menyimpang secara signifikan dari **frekuensi harapan**

Gries (2009: 151)

Chi-Square Test – Goodness-of-fit

- Menguji apakah frekuensi pengamatan untuk kategori dari suatu variabel menyimpang secara signifikan dari frekuensi harapan
- Variabel: ALTERNASI KONSTRUKSIONAL PENEMPATAN PARTIKEL BAHASA INGGRIS
 - Kategori: verb-partikel-objek vs. verb-objek-partikel

Gries (2009: 151)

ALTERNASI KONSTRUKSIONAL PENEMPATAN PARTIKEL BAHASA INGGRIS

Verb-partikel-objek

- He picked up the book.

Gries (2009: 151)

ALTERNASI KONSTRUKSIONAL PENEMPATAN PARTIKEL BAHASA INGGRIS

Verb-partikel-objek

Verb-objek-partikel

- He picked up the book.
- He picked the book up.

Gries (2009: 151)

ALTERNASI KONSTRUKSIONAL PENEMPATAN PARTIKEL BAHASA INGGRIS

Verb-partikel-objek

Verb-objek-partikel

- He **picked up** the book.
- He **picked** the book **up**.

Kedua konstruksi umumnya berterima

Penutur jati seringnya tidak dapat menjelaskan preferensi salah satu dari kedua konstruksi

Gries (2009: 151)

ALTERNASI KONSTRUKSIONAL PENEMPATAN PARTIKEL BAHASA INGGRIS

Verb-partikel-objek

Verb-objek-partikel

- He **picked up** the book.
- He **picked** the book **up**.

Kedua konstruksi umumnya berterima

Penutur jati seringnya tidak dapat menjelaskan preferensi salah satu dari kedua konstruksi

Asumsinya: kedua konstruksi seharusnya sama-sama kerap digunakan

Gries (2009: 151)

ALTERNASI KONSTRUKSIONAL PENEMPATAN PARTIKEL BAHASA INGGRIS

Verb-partikel-objek

Verb-objek-partikel

- He picked up the book.
- He picked the book up.

Gries (2009: 151)

Kedua konstruksi umumnya berterima

Penutur jati seringnya tidak dapat menjelaskan preferensi salah satu dari kedua konstruksi

Asumsinya: kedua konstruksi seharusnya sama-sama kerap digunakan. Dengan kata lain, frekuensi kemunculan kedua konstruksi seharusnya tidak berbeda secara signifikan.

ALTERNASI KONSTRUKSIONAL PENEMPATAN PARTIKEL BAHASA INGGRIS

Verb-partikel-objek

Verb-objek-partikel

- He **picked up** the book.
- He **picked** the book **up**.

Kedua konstruksi umumnya berterima

Penutur jati seringnya tidak dapat menjelaskan preferensi salah satu dari kedua konstruksi

Asumsinya: kedua konstruksi seharusnya sama-sama kerap digunakan. **Dengan kata lain, frekuensi kemunculan kedua konstruksi seharusnya tidak berbeda secara signifikan. <-- KITA AKAN MENGUJI ASUMSI INI**

#1 Merumuskan hipotesis

- Hipotesis Nul (H_0) vs. Hipotesis Alternatif (H_1)
- H_0 :
 - Frekuensi pengamatan kedua tipe/kategori ALTERNASI KONSTRUKSIONAL **identik/setara** – jika ditemukan perbedaan, itu hanyalah perbedaan yang sifatnya acak (*random variation*)
- H_1 :
 - Frekuensi pengamatan kedua tipe/kategori ALTERNASI KONSTRUKSIONAL **tidak** identik/setara, atau berbeda

#1 Merumuskan hipotesis

- Hipotesis Nul (H_0) vs. Hipotesis Alternatif (H_1)
- H_0 :
 - $n_{V \text{ Partikel Obj}} = n_{V \text{ Obj Partikel}}$
- H_1 :
 - $n_{V \text{ Partikel Obj}} \neq n_{V \text{ Obj Partikel}}$

#2 Mengumpulkan data dari korpus

- Memilih korpus
- Mengamati dan menandai kemunculan kedua tipe konstruksi dalam korpus
- Menghitung frekuensi kemunculan kedua tipe konstruksi dalam korpus

#2 Mengumpulkan data dari korpus

Verb-partikel-objek

- He **picked** **up** the book.

Verb-objek-partikel

- He **picked** the book **up**.

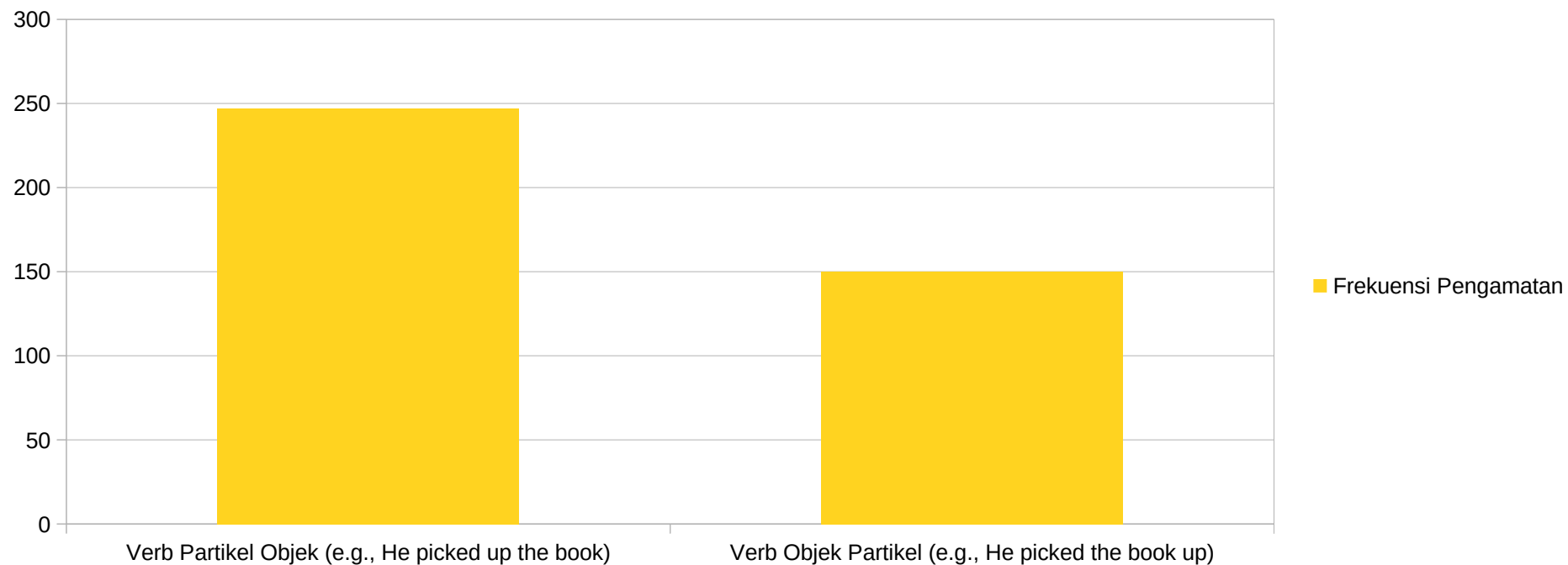
Verb Partikel Objek

247

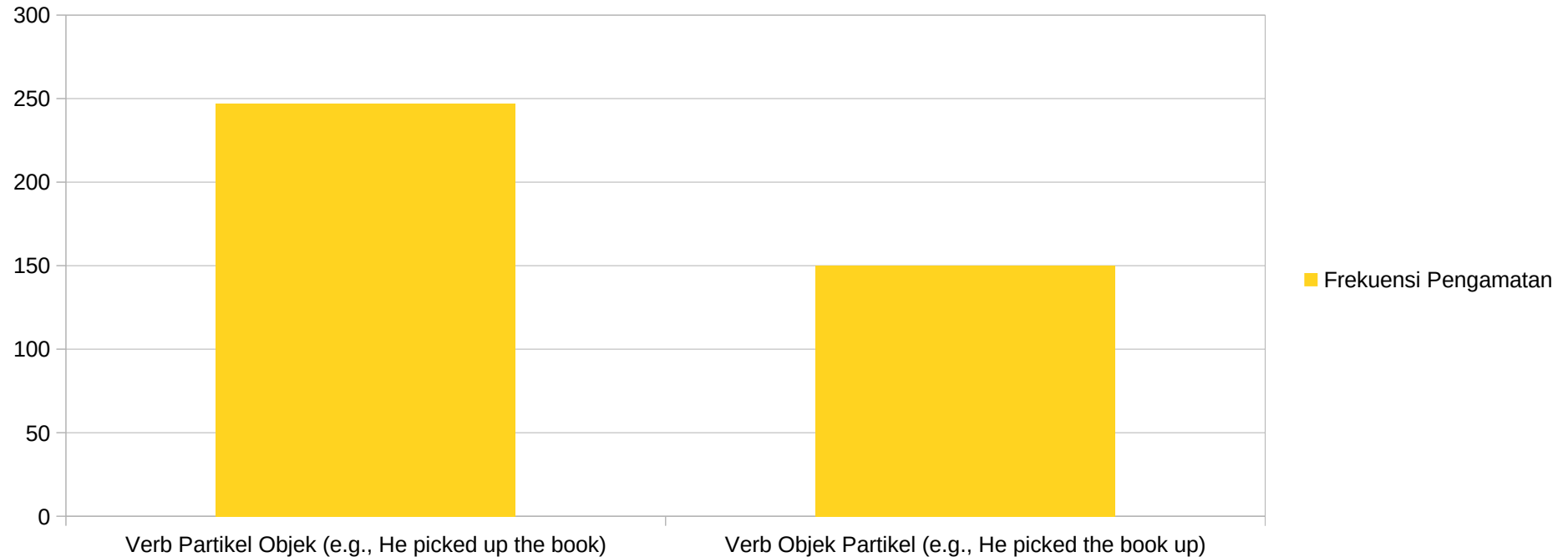
Verb Objek Partikel

150

#2.1 Visualisasi

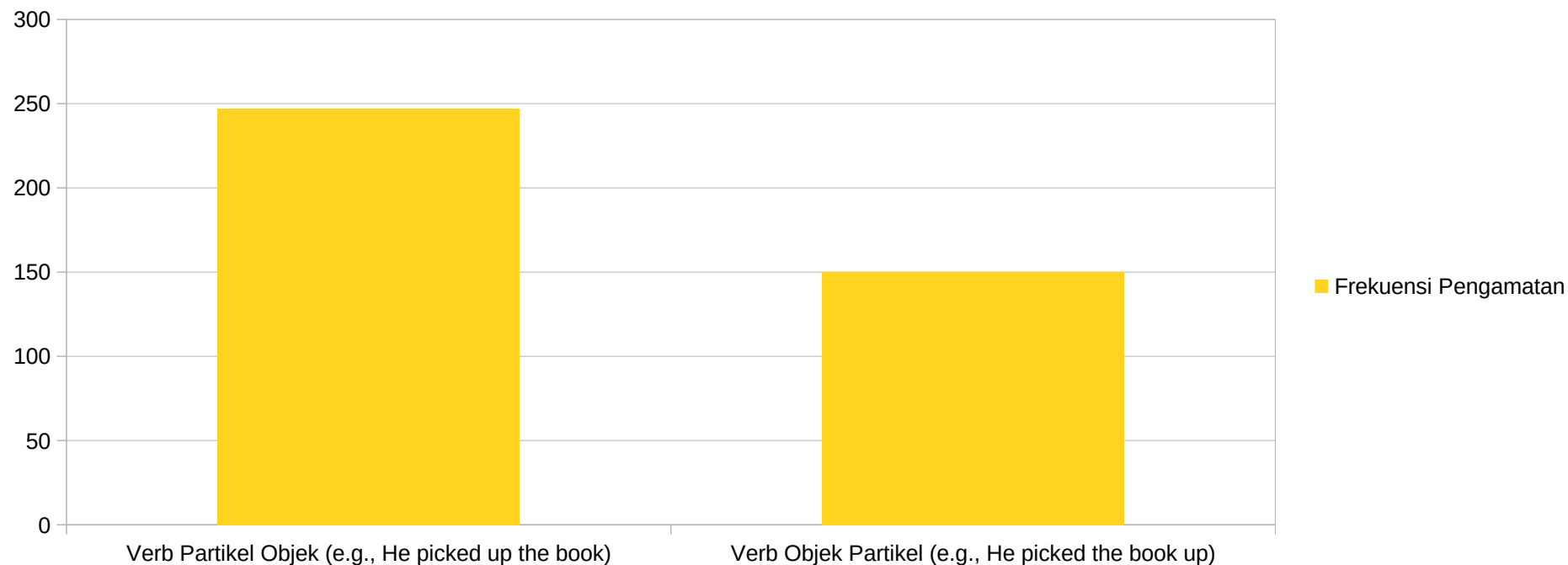


#2.1 Visualisasi



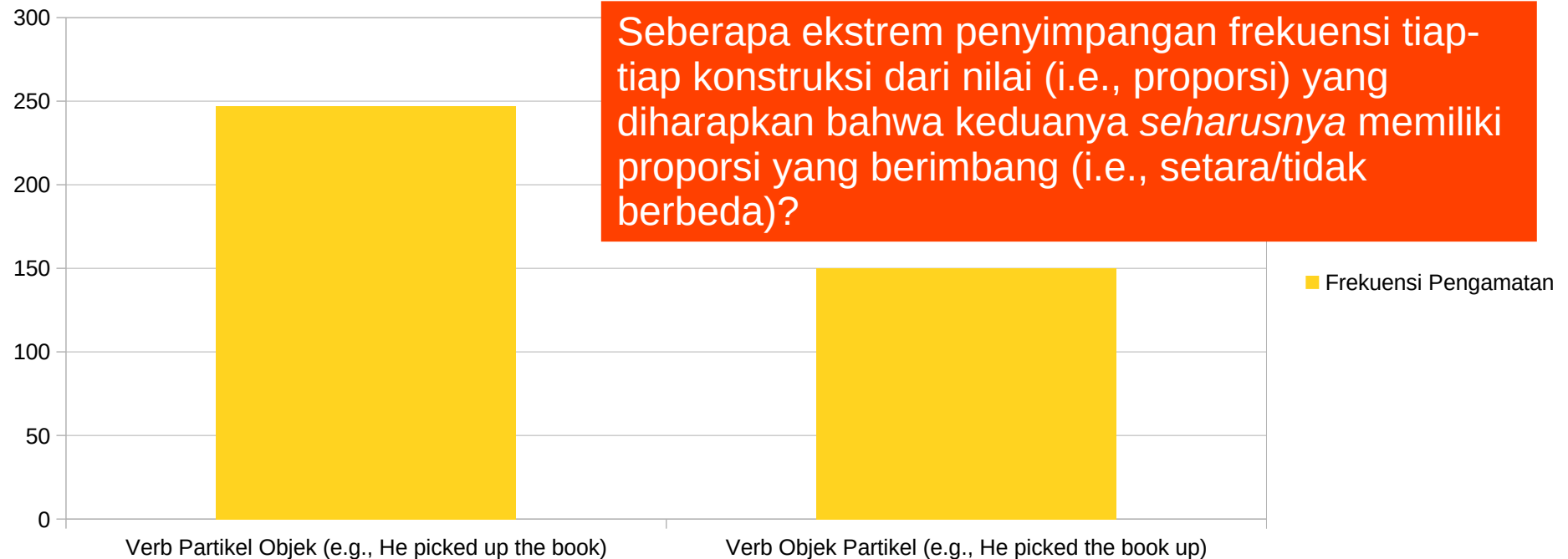
Frekuensi yang diharapkan berdasarkan hipotesis nul adalah frekuensi yang setara.

#2.1 Visualisasi



Frekuensi yang diharapkan berdasarkan hipotesis nul adalah frekuensi yang setara. Artinya bahwa masing-masing tipe konstruksi seharusnya muncul sebanyak 50% (i.e. $\frac{1}{2}$) dari total kemunculan kedua konstruksi dalam sampel.

#2.1 Visualisasi



Frekuensi yang diharapkan berdasarkan hipotesis nul adalah frekuensi yang setara. Artinya bahwa masing-masing tipe konstruksi *seharusnya* muncul sebanyak 50% (i.e. $\frac{1}{2}$) dari total kemunculan kedua konstruksi dalam sampel.

#3 Menghitung frekuensi harapan

- Frekuensi harapan (*expected frequency*) tiap-tiap kategori berdasarkan hipotesis nul adalah frekuensi yang setara
 - Total frekuensi kemunculan kategori (i.e., total frekuensi kemunculan kedua konstruksi) dibagi jumlah kategori (i.e., jumlah tipe konstruksi, yaitu dua)
 - Total frekuensi kategori: $247_{\text{verb-partikel-objek}} + 150_{\text{verb-objek-partikel}} = 397$
 - **Frekuensi harapan** tiap-tiap tipe konstruksi harusnya berimbang/proporsional: $397_{\text{total-frekuensi}} / 2_{\text{tipe konstruksi}} = 198.5$
 - **Kedua konstruksi diharapkan muncul masing-masing sebanyak 198.5 kali** JIKA dianggap kedua tipe konstruksi tersebut seharusnya memiliki distribusi yang sama + kemiripan fungsional secara introspektif.

#3 Menghitung frekuensi harapan

Verb-partikel-objek

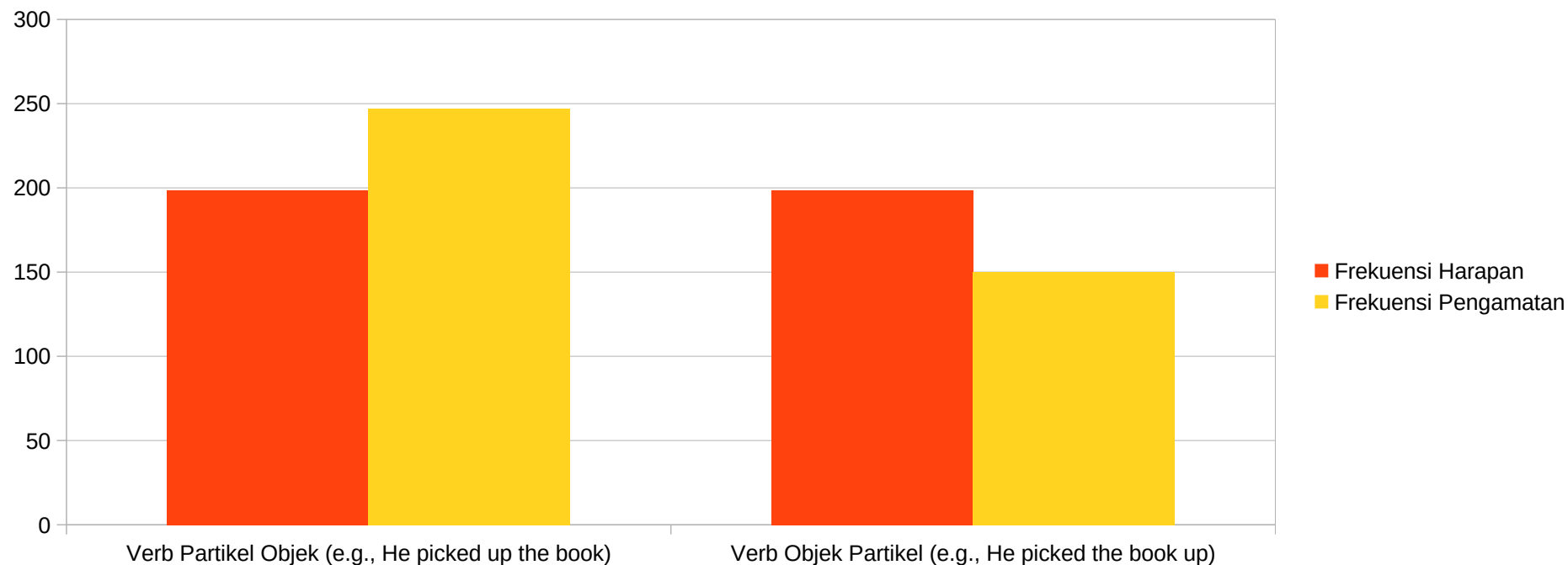
- He **picked** **up** the book.

Verb-objek-partikel

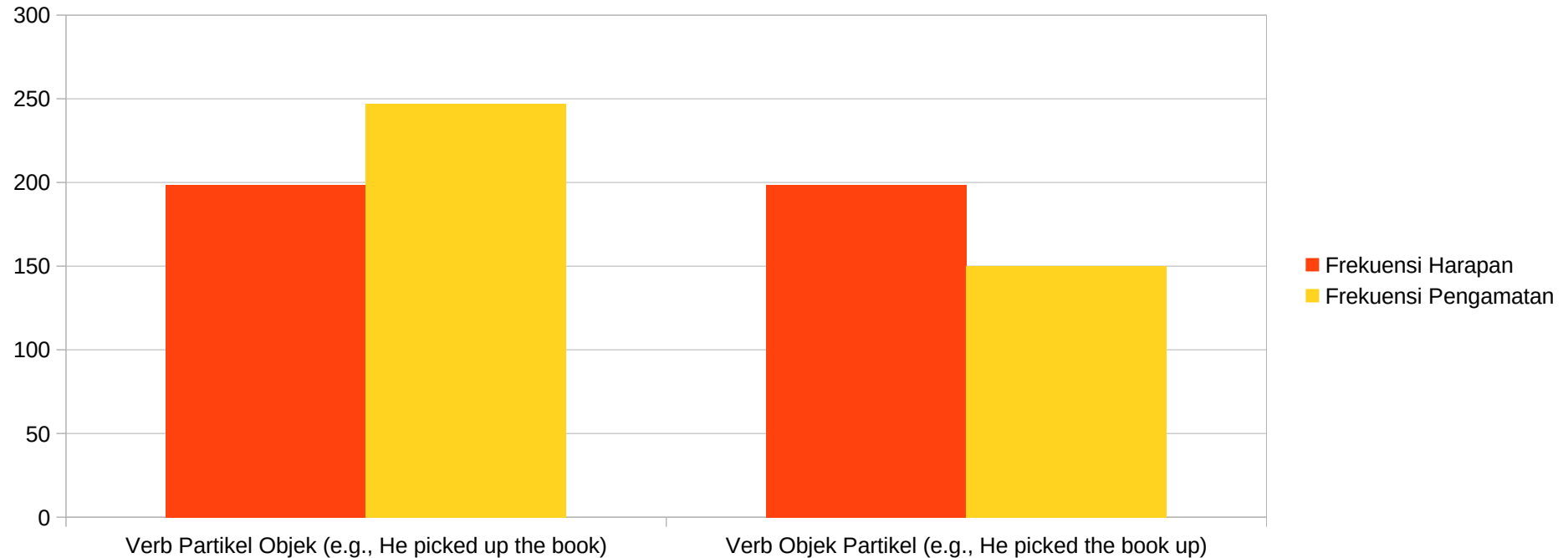
- He **picked** the book **up**.

	Verb Partikel Objek	Verb Objek Partikel	Total
Frekuensi kemunculan/pengamatan (observed freq.)	247	150	397
Frekuensi harapan (expected freq.)	198.5	198.5	397

#3.1 Visualisasi (Obs vs. Exp)



#3.1 Visualisasi (Obs vs. Exp)



Apakah perbedaan frekuensi pengamatan dengan frekuensi harapan adalah suatu penyimpangan yang signifikan secara statistik (yaitu perbedaan yang bukan didasarkan atas kebetulan semata [i.e. *random variation*])?

#4 Uji Chi-Square

- 80% dari tiap-tiap frekuensi harapan harus lebih besar dari atau sama dengan 5
 - Kedua frekuensi harapan pada contoh lebih besar dari 5
 - Kita bisa lakukan uji Chi-Square
 - Jika *p-value* (i.e., *probability of error*) lebih kecil dari 0.05, maka terdapat perbedaan distribusi yang signifikan secara statistik antara frekuensi yang diamati dalam data dengan yang diharapkan
 - In other words, ketika $p\text{-value} < 0.05$, terdapat peluang yang sangat kecil (i.e., lebih kecil dari 5%) untuk mengamati nilai distribusi dalam sampel JIKA memang benar adanya (menurut H_0) bahwa konstruksi VPO dan VOP seharusnya memiliki distribusi yang setara dalam populasi.

Google Sheet Demo

Demo di R