

## Indexing ICE-GB and ICECUP

- ◆
  - Spoken and written: 60% spoken • 500 2,000-word texts = 1Mw
- ◆
  - Structural markup, tagging and parsing (based on Quirk 1985)

## Query

- ◆
  - Retrieval levels of browsing: - overview - text - sentence
  - Search by sociolinguistic variable, text string or FF
- ◆
  - An intuitively model-based grammatical query system
- ◆
  - Sufficiently expressive for a large range of experiments
  - Ask questions which could not consider before
  - No programming required...
  - ...but we still have to think...

---

# S a i i and xp im n al d ign



- A: o g n ralis vid nc from a corpus to "R al Languag "



- $e$  ,  $e$  .
- A hypothesis consists of an
  - independent variable ( )
  - dependent variable ( )
- i . Do s t valu of t  $V_r$  av an ff ct on t valu of t DV?
- $e$  = t pr diction t at t r is ho ff ct.



- Q s "w om" us d mor oft n t an "w o" in
-

---

## S a i i and xp im n al d ign (II)



• **e e q e** can tell you how common a word is in the corpus. But the reason that it is there might depend on many irrelevant factors.

• **e e e q e** focuses on variation where there is a choice. It tells you how often the speaker or writer chooses to use one word over another. It lets us focus on a specific type of **e e**.



*Ca u L, u, c*

□ n - lid xp im n g id

- ◆
- ◆
- ◆
- ◆

	d, p, nd, nt variabl			
		, ^	,	

- ◆

• for xampl (abov ) do sp ak rs positiv ly f oos ?



# Performing a chi-squared test

## ◆ $\chi^2$

- cf. observed vs expected distributions:
- Simplified, specific value of DV: on obs. column (e.g. )
  - Observed **O** = specific value of DV
  - Expected **E** = total value of DV, scaled down
- OR all values of DV: sum all columns
- Formula:

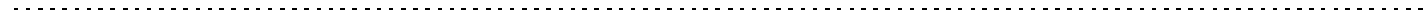
- Test: is it significant that an attribute's old value  $\chi^2$ ?

## ◆

- degrees of freedom  $\chi^2$ 
  - or  $w - r$

.....







---

## Exercise 2: gamma $\Rightarrow$ gamma



- Does the 'mood' of a clause predict its transitivity?
- How does one element within a clause or phrase affect another?



- We must specify the (g. t. clause or phrase)
  - We're allowed to consider unmarked cases, g. d i
-

