Including ICE–GB and ICECUP

♦
- Spoken and written: 60% spoken • 500 2,000-word texts = 1Mw

♦
- Structural markup, tagging and parsing (based on Quirk 1985)

Utility

♦
- Three levels of browsing: - overview - text - sentence
- Search by sociolinguistic variable, text string or FTF

♦
- An intuitive model-based grammatical query system

♦
- Sufficiently expressive for a huge range of experiments
- Ask questions we could not consider before
- No programming required...
- ...but we still have to think...

ipcw

Statistics and Experimental Design

- ♦
  - A: to generalise evidence from a corpus to "Real Language"
- ♦

  - .
  - A hypothesis consists of an
    - independent variable (IV)
    - dependent variable (DV)
    - i. Does the value of the IV have an effect on the value of the DV?
    - = the prediction that there is no effect.
- ♦

  - Q: Is "whom" used more often than "who" in

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

◆

- . can t ll you ow common a word is in t corpus. But t r ason t at it is t r mig t d p nd on many irr l vant factors.
- . focus s on variation w r t r is a c oic . t t lls you ow oft n t sp ak r or writ r c oos s to us on word ov r anot r. t l ts us focus on a sp cific typ of .

◆

■ ■n – lid ■xp ■im■n g id

- ◆
- ◆
- ◆
- ◆

		d■p■nd■nt variabl■			
			,		
	,	, ^	,		

- ◆

- for ■ xampl■ (abov■) do sp■ak■rs positiv■ ly ꞔ oos■ ?

P f ming a a i i al

- ♦ χ
  - cf. obs rv d vs xp ct d distributions:
  - Simpl , sp cific valu of DV: on obs. column (.g.)
    - Obs rv d **O** = sp cific valu of DV
    - Exp ct d **E** = total valu of DV, scal d down
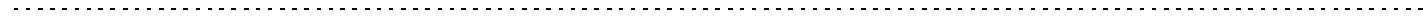 - OR all valu s of DV: sum all columns
  - Formula:

  - st: is t is gr at r t an a t r s old valu χ ?
- ♦ χ
  - d gr s of fr dom
    - or w r

Ex i 2: g amma ⇒ g amma

♦

- Do s t 'mood' of a claus pr dict its transitivity?
- How do s on l m nt wit in a claus or p ras aff ct anot r?

♦

- W must sp cify t (g. t claus or p ras)
- W r av to consid r unmark d cas s, g. d i