Detecting direction in interaction evidence

Sean Wallis, Survey of English Usage, University College London
April 2017

1. Introduction

I have previously argued (Wallis 2014) that interaction evidence is the most fruitful type of corpus linguistics evidence for grammatical research (and doubtless for many other areas of linguistics).

Frequency evidence, which we can write as ***p***(***x***), the probability of ***x*** occurring, concerns itself simply with the overall distribution of linguistic phenomenon ***x*** – such as whether informal written English has a higher proportion of interrogative clauses than formal written English. In order to calculate frequency evidence we must define ***x***, i.e. decide how to identify interrogative clauses. We must also pick an appropriate baseline ***n*** for this evaluation, i.e. we need to decide whether to use words, clauses, or any other structure to identify locations where an interrogative clause may occur.

Interaction evidence is different. It is a statistical correlation between a decision that a writer or speaker makes at one part of a text, which we will label point ***A***, and a decision at another part, point ***B***. The idea is shown schematically in Figure 1. ***A*** and ***B*** are separate 'decision points' in a given relationship (e.g. lexical adjacency), which can be also considered as 'variables'.

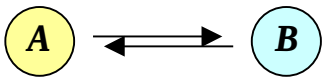

Figure 1: Associative inference from lexico-grammatical choice variable ***A*** to variable ***B*** (sketch).

This class of evidence is used in a wide range of computational algorithms. These include collocation methods, part-of-speech taggers, and probabilistic parsers. Despite the promise of interaction evidence, however, the majority of theory-driven corpus studies tend to consist of discussions of frequency differences and distributions.

In this paper I want to consider applications of interaction evidence which are made more-or-less at the same time by the same speaker/writer. In such circumstances we cannot be sure that just because ***B*** follows ***A*** in the text, the decision relating to ***B*** was made after the decision at ***A***.

For example, in studying the premodification of noun phrases by attributive adjectives in English – which adjective is applied first in assembling an NP like ***the old tall green ship***, for instance – ***we cannot be sure that the adjectives are selected by the speaker in sentence order***. It is also perfectly plausible that they were sh at ld in e in aite igati5585(y2.063 12 Tf 7.32585(e)3.74( )-7.80439(
nr theel iniy

Copyright © Survey of English Usage, UCL

Howsoever desirable it may be, detecting decision-making order by ***post hoc*** stochastic methods is not actually possible. What this paper does is investigate methods for determining ***the relative size of effect*** of one variable on another and vice versa.

A and ***B*** may interact, but some interactions are one-sided, i.e. directional.

2. A collocation example

Let us consider a simple example 'experiment' whose result we can predict. In British English, LOOK ***askance*** is an archaic idiom. The adverb ***askance*** almost never appears without being preceded by the lemma LOOK.

However – and here is the power of an intuitive example – ***the reverse is not true***. The most common words that follow LOOK are prepositions ***at*** (26,629) and ***for*** (8,117). Among adverbs, LOOK ***back***/***forward*** (at 2,170 and 2,518 respectively) or ***up***/***down*** (3,634; 2,167) are far more frequent than the rare LOOK ***askance***.

The question is how we can estimate the 'one-sidedness' of the relationship between LOOK and ***askance***? Using Mark Davies' interface to the ***British National Corpus*** (BNC), we obtain the statistics in Table 1.

	Frequency	Probability
LOOK	105,871	0.00105871
askance	48	0.00000048
LOOK ***askance***	31	0.00000031
p(LOOK) × ***p***(***askance***)		5.0818×10^{-10}
p(LOOK | ***askance***) = 31/48		0.64583333
p(***askance*** | LOOK) = 31/105,871		0.00029281

Table 1: Sample frequency data for LOOK, ***askance***, and LOOK ***askance*** from the BNC[1] and some derived probabilities.

The notation '***p***(LOOK | ***askance***)' means the probability of the verb lemma LOOK being uttered if the following word is ***askance***.

- The probability of the word ***askance*** being uttered in the corpus is 48 in 100,000,000 words, or 0.00000048 (or 0.000048% if you prefer). But if the previous word is LOOK, that probability is

Copyright © Survey of English Usage, UCL

These per million word statistics are exposure statistics.

- If we hear LOOK, the probability of the next word being ***askance*** increases by around 0.03%. Although the probability increases, this low overall probability would not cause us to 'expect' it. LOOK ***at*** or LOOK ***for*** is far more likely (33% compared to 0.03%).

- But if we misheard the verb, and then heard the word ***askance***, the chance of the previous word being ***look***, ***looks***, ***looked***, or ***looking***

Copyright © Survey of English Usage, UCL

	LOOK	**¬LOOK**	**total**

Copyright © Survey of English Usage, UCL

Detect65558(e)3.740 1 56.6395 795.801 Tm [(D)1.57503(e)3.74(t)-2.16558(e)3.74(c)3.74(t)-2.16558(65558(e)3

3.1 Testing for direction under alternation

How do we test for directionality when a variable can freely vary from 0 to 1 and where both values (negative and positive) should be considered?

In our first example, we used a 2×2 χ^2 test for homogeneity (association) to compare the two variables $\boldsymbol{A}$ = {LOOK, ¬LOOK}, $\boldsymbol{B}$ = {***askance***, ¬***askance***}. The test compares both values of each variable, i.e. {$\boldsymbol{a}$, ¬$\boldsymbol{a}$} × {$\boldsymbol{b}$. ¬$\boldsymbol{b}$}.

We then used goodness of fit tests to examine the changing probability of selecting a word. Our method compared $\boldsymbol{d}_1 \neq \boldsymbol{d}_2$ where

$\boldsymbol{d}_1 = \boldsymbol{p}(\boldsymbol{b} \mid \boldsymbol{a}) - \boldsymbol{p}(\boldsymbol{b}) = \boldsymbol{p}(\boldsymbol{askance} \mid \text{LOOK}) - \boldsymbol{p}(\boldsymbol{askance})$, and
$\boldsymbol{d}_2 = \boldsymbol{p}(\boldsymbol{a} \mid \boldsymbol{b}) - \boldsymbol{p}(\boldsymbol{a}) = \boldsymbol{p}(\text{LOOK} \mid \boldsymbol{askance}) - \boldsymbol{p}(\text{LOOK})$.

In this second test we only tested one value of both variables – the chance of selecting the word, $\boldsymbol{p}(\boldsymbol{a})$ and $\boldsymbol{p}(\boldsymbol{b})$. We did not consider the chance of selecting any other word**.** We did not compare, for instance, $\boldsymbol{p}(\neg\boldsymbol{a})$ and $\boldsymbol{p}(\boldsymbol{b})$:

$\boldsymbol{d}_1 = \boldsymbol{p}(\boldsymbol{b} \mid \neg\boldsymbol{a}) - \boldsymbol{p}(\boldsymbol{b}) = \boldsymbol{p}(\boldsymbol{askance} \mid \neg\text{LOOK}) - \boldsymbol{p}(\boldsymbol{askance})$, and
$\boldsymbol{d}_2 = \boldsymbol{p}(\neg\boldsymbol{a} \mid \boldsymbol{b}) - \boldsymbol{p}(\neg\boldsymbol{a}) = \boldsymbol{p}(\neg\text{LOOK} \mid \boldsymbol{askance}) - \boldsymbol{p}(\neg\text{LOOK})$.

This seems intuitive in this case: surely this doesn't matter – all values of $\boldsymbol{p}(\neg\boldsymbol{word})$ except $\boldsymbol{p}(\neg\text{LOOK} \mid \boldsymbol{askance})$ are likely to be close to 1! The fact that we don't even consider this prospect is probably a consequence of the fact that these values are not freely alternating.

If we employ this method in the grammatical example, however, we get four distinct results for each combination {$\boldsymbol{a}$, ¬$\boldsymbol{a}$} × {$\boldsymbol{b}$. ¬$\boldsymbol{b}$}. This is summarised visually by Figure 7.

We are not weighting all cells in the contingency table equally. We obtain different results depending on which we pick. Three out of four represent a significant difference, and one is not significant. Which should we choose?

3.2 Comparing Newcombe-Wilson intervals for direction

Instead of comparing goodness of fit tests, we propose to compare two Newcombe-Wilson tests (Wallis 2013b). This method tests if $\boldsymbol{d}_1 \neq \boldsymbol{d}_2$ where

$\boldsymbol{d}_1 = \boldsymbol{p}(\boldsymbol{b} \mid \neg\boldsymbol{a}) - \boldsymbol{p}(\boldsymbol{b} \mid \boldsymbol{a})$, and
$\boldsymbol{d}_2 = \boldsymbol{p}(\boldsymbol{a} \mid \neg\boldsymbol{b}) - \boldsymbol{p}(\boldsymbol{a} \mid \boldsymbol{b})$.

Table 3b summarises the paired test for homogeneity.[7] Individual Newcombe-Wilson tests are significant, that is, we can report that the polarity of the question tag affects the polarity of the VP (TEST 1), and the decision of whether to employ a positive or negative VP has an effect on the polarity of the question tag (TEST 2).

[7] To achieve this import data into the spreadsheet for testing separability between two 2×2 homogeneity tests (select the '2x2 homogeneity' tab).

Copyright © Survey of English Usage, UCL

equally – with the result that there is no significant difference in direction. As a result, the direction test contour looks very different than that for association.

5. Concluding remarks

This evaluation performs three different significance tests, one of which is employed twice.

- The 2×2 ^{2} test simply tests for **association**, i.e. whether the two variables (outcomes of choices at ***A*** and ***B***) interact.
- The second test (the 2×1 goodness of fit ^{2} test or the Newcombe-Wilson test) obtains **directional** information. It is used twice:
  - to test whether making a particular choice at ***A*** correlates with the chance of making a particular choice at ***B*** to significantly increase; and
  - to test whether making a choice at point ***B*** correlates with an increased propensity to make a particular choice at point ***A***.
- These two tests are then contrasted with a **separability test**. This evaluates whether the increase in one direction is significantly greater than in the other.

The first example we used above was based on a simp

Copyright © Survey of English Usage, UCL