Empirical methods in phonological research

Michael Hammond

U. of Arizona

Outline

Data collection in phonology

Data collection in phonology

1. Grammaticality judgments

Data collection in phonology

1. Grammaticality judgments
2. Fieldwork

Data collection in phonology

1. Grammaticality judgments
2. Fieldwork
3. Experiments

Data collection in phonology

1. Grammaticality judgments
2. Fieldwork
3. Experiments
4. Corpus studies

Data collection in phonology

1. Grammaticality judgments
2. Fieldwork
3. Experiments
4. Corpus studies
5. Computational simulations

Data collection in phonology

1. Grammaticality judgments
2. Fieldwork
3. Experiments
4. Corpus studies
5. Computational simulations
6. Making it up. . .

Data collection in phonology

1. Grammaticality judgments
2. Fieldwork
3. Experiments
4. Corpus studies
5. Computational simulations
6. Making it up. . .

Working with corpora

Working with corpora

- Kinds of corpora

Working with corpora

- Kinds of corpora
- Obtaining a corpus

Working with corpora

- Kinds of corpora
- Obtaining a corpus
- Tools

Working with corpora

- Kinds of corpora
- Obtaining a corpus
- Tools
- Statistics

Working with corpora

- Kinds of corpora
- Obtaining a corpus
- Tools
- Statistics
- Practicality

Kinds of corpora

Kinds of corpora

- Written texts, e.g. BNC, Brown, WSJ, Gutenberg, etc.

Kinds of corpora

- Written texts, e.g. BNC, Brown, WSJ, Gutenberg, etc.
- Treebanks, e.g. Suzanne, etc.

Kinds of corpora

- Written texts, e.g. BNC, Brown, WSJ, Gutenberg, etc.
- Treebanks, e.g. Suzanne, etc.
- Dictionaries, e.g. CMU, Moby, MRC, etc.

Kinds of corpora

- Written texts, e.g. BNC, Brown, WSJ, Gutenberg, etc.
- Treebanks, e.g. Suzanne, etc.
- Dictionaries, e.g. CMU, Moby, MRC, etc.
- Acoustic data, e.g. TIMIT, Callhome, etc.

Kinds of corpora

- Written texts, e.g. BNC, Brown, WSJ, Gutenberg, etc.
- Treebanks, e.g. Suzanne, etc.
- Dictionaries, e.g. CMU, Moby, MRC, etc.
- Acoustic data, e.g. TIMIT, Callhome, etc.
- The web itself

Kinds of corpora

- Written texts, e.g. BNC, Brown, WSJ, Gutenberg, etc.
- Treebanks, e.g. Suzanne, etc.
- Dictionaries, e.g. CMU, Moby, MRC, etc.
- Acoustic data, e.g. TIMIT, Callhome, etc.
- The web itself

Most of these are less than optimal for phonological work. They are mostly written and typically not transcribed.

Obtaining a corpus

Obtaining a corpus

This is probably the hardest part of working with corpus data.

Obtaining a corpus

This is probably the hardest part of working with corpus data.

- Determining what is optimal, yet obtaining something that is workable.

Obtaining a corpus

This is probably the hardest part of working with corpus data.

- Determining what is optimal, yet obtaining something that is workable.
- What is in the public domain?

Obtaining a corpus

This is probably the hardest part of working with corpus data.

- Determining what is optimal, yet obtaining something that is workable.
- What is in the public domain?
- What does it cost? What can be gotten for free?

Tools

Tools

- The format of a corpus is often a challenge.

Tools

- The format of a corpus is often a challenge.
  - No transcription.

Tools

- The format of a corpus is often a challenge.
  - No transcription.
  - Odd transcription.

Tools

- The format of a corpus is often a challenge.
  - No transcription.
  - Odd transcription.
  - No glosses.

Tools

- The format of a corpus is often a challenge.
  - No transcription.
  - Odd transcription.
  - No glosses.
  - Odd glosses.

Tools

- The format of a corpus is often a challenge.
  - No transcription.
  - Odd transcription.
  - No glosses.
  - Odd glosses.
  - Normalization issues.

Tools

- The format of a corpus is often a challenge.
  - No transcription.
  - Odd transcription.
  - No glosses.
  - Odd glosses.
  - Normalization issues.
  - Odd extra information.

Tools

- The format of a corpus is often a challenge.
  - No transcription.
  - Odd transcription.
  - No glosses.
  - Odd glosses.
  - Normalization issues.
  - Odd extra information.
- Tools for searching and counting

Tools

- The format of a corpus is often a challenge.
  - No transcription.
  - Odd transcription.
  - No glosses.
  - Odd glosses.
  - Normalization issues.
  - Odd extra information.
- Tools for searching and counting
  - General easy commercial tools, e.g. Monoconc, Shoebox, MSWord, etc.

Tools

- The format of a corpus is often a challenge.
  - No transcription.
  - Odd transcription.
  - No glosses.
  - Odd glosses.
  - Normalization issues.
  - Odd extra information.
- Tools for searching and counting
  - General easy commercial tools, e.g. Monoconc, Shoebox, MSWord, etc.
  - Generic search tools, e.g. `grep`, `findstr`, etc.

Tools

- The format of a corpus is often a challenge.
  - No transcription.
  - Odd transcription.
  - No glosses.
  - Odd glosses.
  - Normalization issues.
  - Odd extra information.
- Tools for searching and counting
  - General easy commercial tools, e.g. Monoconc, Shoebox, MSWord, etc.
  - Generic search tools, e.g. `grep`, `findstr`, etc.
  - Programming tools, e.g. Perl, Awk, etc.

Statistics

Statistics

How do you know that you've found something that is true?

Statistics

How do you know that you've found something that is true?

- It depends on your claim:

Statistics

How do you know that you've found something that is true?

- It depends on your claim:
  - X occurs.

Statistics

How do you know that you've found something that is true?

- It depends on your claim:
  - X occurs.
  - X occurs a lot.

Statistics

How do you know that you've found something that is true?

- It depends on your claim:
    - X occurs.
    - X occurs a lot.
    - X occurs more than Y.

Statistics

How do you know that you've found something that is true?

- It depends on your claim:
  - X occurs.
  - X occurs a lot.
  - X occurs more than Y.
- It depends on the discipline:

Statistics

How do you know that you've found something that is true?

- It depends on your claim:
  - X occurs.
  - X occurs a lot.
  - X occurs more than Y.
- It depends on the discipline:
  - Psycholinguistics.

Statistics

How do you know that you've found something that is true?

- It depends on your claim:
  - X occurs.
  - X occurs a lot.
  - X occurs more than Y.
- It depends on the discipline:
  - Psycholinguistics.
  - Computational linguistics.

Statistics

How do you know that you've found something that is true?

- It depends on your claim:
  - X occurs.
  - X occurs a lot.
  - X occurs more than Y.
- It depends on the discipline:
  - Psycholinguistics.
  - Computational linguistics.
  - Phonology.

Practicality

Practicality

One can't overstate the importance of solving the practical issues of corpus work to actually do something of interest.

Practicality

One can't overstate the importance of solving the practical issues of corpus work to actually do something of interest.

- What kind of corpus can you get your hands on?

Practicality

One can't overstate the importance of solving the practical issues of corpus work to actually do something of interest.

- What kind of corpus can you get your hands on?
- Is it appropriate for your goals?

Practicality

One can't overstate the importance of solving the practical issues of corpus work to actually do something of interest.

- What kind of corpus can you get your hands on?
- Is it appropriate for your goals?
- Do you have appropriate tools to find what you want?

Practicality

One can't overstate the importance of solving the practical issues of corpus work to actually do something of interest.

- What kind of corpus can you get your hands on?
- Is it appropriate for your goals?
- Do you have appropriate tools to find what you want?
- How do you know you've found something?

A simple example

A simple example

- Hypothesis: Words that begin with voiced oral stops are more frequent than words that begin with nasal stops in English.

A simple example

- Hypothesis: Words that begin with voiced oral stops are more frequent than words that begin with nasal stops in English.
- Let's make some commitment about corpus and tools.

A simple example

- Hypothesis: Words that begin with voiced oral stops are more frequent than words that begin with nasal stops in English.
- Let's make some commitment about corpus and tools.
  - Brown corpus for data

A simple example

- Hypothesis: Words that begin with voiced oral stops are more frequent than words that begin with nasal stops in English.
- Let's make some commitment about corpus and tools.
  - Brown corpus for data
  - `findstr`/`grep` to do the counts

A simple example

- Hypothesis: Words that begin with voiced oral stops are more frequent than words that begin with nasal stops in English.
- Let's make some commitment about corpus and tools.
  - Brown corpus for data
  - `findstr`/`grep` to do the counts
  - `R` for statistics

A simple example

- Hypothesis: Words that begin with voiced oral stops are more frequent than words that begin with nasal stops in English.
- Let's make some commitment about corpus and tools.
  - Brown corpus for data
  - `findstr`/`grep` to do the counts
  - `R` for statistics
- These choices are for pedagogical purposes.

The challenges

The challenges

- Brown is an orthographic corpus.

The challenges

- Brown is an orthographic corpus.
- `grep` and especially `findstr` are rather weak tools.

The challenges

- Brown is an orthographic corpus.
- `grep` and especially `findstr` are rather weak tools.
- What does "more frequent than" mean?

What do we count?

What do we count?

- It would seem we should compare [b,d,g] with [m,n,ŋ].

What do we count?

- It would seem we should compare [b,d,g] with [m,n,ŋ].
- However, we have to throw out [ŋ] because it doesn't occur word-initially in English.

What do we count?

- It would seem we should compare [b,d,g] with [m,n,ŋ].
- However, we have to throw out [ŋ] because it doesn't occur word-initially in English.
- We should then also throw out [g] because we want the same number of "categories".

What do we count?

- It would seem we should compare [b,d,g] with [m,n,ŋ].
- However, we have to throw out [ŋ] because it doesn't occur word-initially in English.
- We should then also throw out [g] because we want the same number of "categories".
- This is rather convenient because <g> maps to both [g] and [ǰ].

What do we count?

- It would seem we should compare [b,d,g] with [m,n,ŋ].
- However, we have to throw out [ŋ] because it doesn't occur word-initially in English.
- We should then also throw out [g] because we want the same number of "categories".
- This is rather convenient because <g> maps to both [g] and [ǰ].
- We've chosen sounds that map relatively directly to English spelling, but we still have to deal with some oddities, e.g. words like *know* [no].

How do we count?

How do we count?

- Under Unix or Mac, use `grep` and `wc`.

How do we count?

- Under Unix or Mac, use `grep` and `wc`.
  - `grep -Eo '\<(kn|Kn|[mnMN])[a-z]*\>' brown.txt`
  - `grep -Eo '\<[bdBD][a-z]*\>' brown.txt`

How do we count?

- Under Unix or Mac, use `grep` and `wc`.
  - `grep -Eo '\<(kn|Kn|[mnMN])[a-z]*\>' brown.txt`
  - `grep -Eo '\<[bdBD][a-z]*\>' brown.txt`
  - Pipe to `wc` and count lines: `grep pattern file | wc`

How do we count?

- Under Unix or Mac, use `grep` and `wc`.
  - `grep -Eo '\<(kn|Kn|[mnMN])[a-z]*\>' brown.txt`
  - `grep -Eo '\<[bdBD][a-z]*\>' brown.txt`
  - Pipe to `wc` and count lines: `grep pattern file | wc`
- Under Windows, it's a little trickier.

How do we count?

- Under Unix or Mac, use `grep` and `wc`.
  - `grep -Eo '\<(kn|Kn|[mnMN])[a-z]*\>' brown.txt`
  - `grep -Eo '\<[bdBD][a-z]*\>' brown.txt`
  - Pipe to `wc` and count lines: `grep pattern file | wc`
- Under Windows, it's a little trickier.
  - Use MSWord to massage the browncorpus.txt file into something searchable. Use the global search-and-replace function to put each word on a separate line by replacing spaces with a return.

How do we count?

- Under Unix or Mac, use `grep` and `wc`.
  - `grep -Eo '\<(kn|Kn|[mnMN])[a-z]*\>' brown.txt`
  - `grep -Eo '\<[bdBD][a-z]*\>' brown.txt`
  - Pipe to `wc` and count lines: `grep pattern file | wc`
- Under Windows, it's a little trickier.
  - Use MSWord to massage the browncorpus.txt file into something searchable. Use the global search-and-replace function to put each word on a separate line by replacing spaces with a return.
  - Use `findstr` in the DOS window to find lines that match the pattern, and pipe the results to a text file:

How do we count?

- Under Unix or Mac, use `grep` and `wc`.
  - `grep -Eo '\<(kn|Kn|[mnMN])[a-z]*\>' brown.txt`
  - `grep -Eo '\<[bdBD][a-z]*\>' brown.txt`
  - Pipe to `wc` and count lines: `grep pattern file | wc`
- Under Windows, it's a little trickier.
  - Use MSWord to massage the browncorpus.txt file into something searchable. Use the global search-and-replace function to put each word on a separate line by replacing spaces with a return.
  - Use `findstr` in the DOS window to find lines that match the pattern, and pipe the results to a text file:
  - `findstr "^[bdBD]" bc.txt > bdwords.txt`

How do we count?

- Under Unix or Mac, use `grep` and `wc`.
  - `grep -Eo '\<(kn|Kn|[mnMN])[a-z]*\>' brown.txt`
  - `grep -Eo '\<[bdBD][a-z]*\>' brown.txt`
  - Pipe to `wc` and count lines: `grep pattern file | wc`
- Under Windows, it's a little trickier.
  - Use MSWord to massage the browncorpus.txt file into something searchable. Use the global search-and-replace function to put each word on a separate line by replacing spaces with a return.
  - Use `findstr` in the DOS window to find lines that match the pattern, and pipe the results to a text file:
  - `findstr "^[bdBD]" bc.txt > bdwords.txt`
  - `findstr "^kn ^Kn ^[mnMN]" bc.txt > mnwords.txt`

How do we count?

- Under Unix or Mac, use `grep` and `wc`.
  - `grep -Eo '\<(kn|Kn|[mnMN])[a-z]*\>' brown.txt`
  - `grep -Eo '\<[bdBD][a-z]*\>' brown.txt`
  - Pipe to `wc` and count lines: `grep pattern file | wc`
- Under Windows, it's a little trickier.
  - Use MSWord to massage the browncorpus.txt file into something searchable. Use the global search-and-replace function to put each word on a separate line by replacing spaces with a return.
  - Use `findstr` in the DOS window to find lines that match the pattern, and pipe the results to a text file:
  - `findstr "^[bdBD]" bc.txt > bdwords.txt`
  - `findstr "^kn ^Kn ^[mnMN]" bc.txt > mnwords.txt`
  - Open the resulting text file in `Notepad`, making sure the 'status line' is displayed, and move the mouse to the last line of the file.

Challenges

Challenges

There are a number of challenges to doing this kind of work.

Challenges

There are a number of challenges to doing this kind of work.

- Command-line interface

Challenges

There are a number of challenges to doing this kind of work.

- Command-line interface
- Search expressions (regular expression syntax)

Challenges

There are a number of challenges to doing this kind of work.

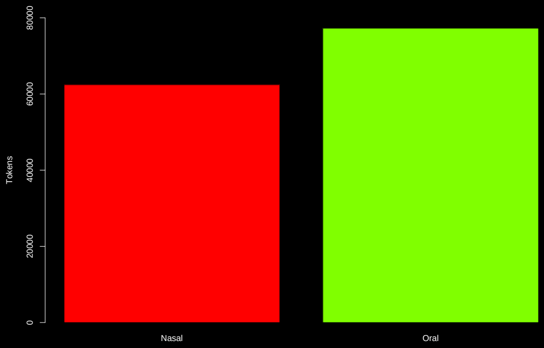
- Command-line interface
- Search expressions (regular expression syntax)
- Tool chains (piping and redirection)

Results

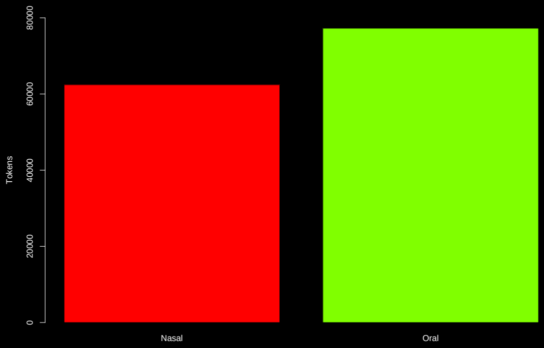
total number of words	1,026,595
words beginning with [mn]	62,415
words beginning with [bd]	77,262

What does this look like?

What does this look like?



What does this look like?




```
> barplot(c(62415,77262))
```

Is the hypothesis correct?

How do we know the difference between 62,415 and 77,262 isn't due to chance?

Is the hypothesis correct?

How do we know the difference between 62,415 and 77,262 isn't due to chance? χ^2 (chi-square).

Is the hypothesis correct?

How do we know the difference between 62,415 and 77,262 isn't due to chance? χ^2 (chi-square).

- ```
  > chisq.test(c(62415,77262))

          Chi-squared test for given probabilities

  data:  c(62415, 77262)
  X-squared = 1578.165, df = 1, p-value < 2.2e-16
  ```
  ```

Is the hypothesis correct?

How do we know the difference between 62,415 and 77,262 isn't due to chance? χ^2 (chi-square).

-
```
> chisq.test(c(62415,77262))

        Chi-squared test for given probabilities

data:  c(62415, 77262)
X-squared = 1578.165, df = 1, p-value < 2.2e-16
```
- *p* is really small, much smaller than .05; our hypothesis is correct.

Another corpus question

Another corpus question

- Is there a difference in the number of types (distinct words) that begin with voiced stops vs. nasals?

Another corpus question

- Is there a difference in the number of types (distinct words) that begin with voiced stops vs. nasals?
- This can be tested more easily with a dictionary, rather than a text.

Another corpus question

- Is there a difference in the number of types (distinct words) that begin with voiced stops vs. nasals?
- This can be tested more easily with a dictionary, rather than a text.
- For Unix geeks, there is another way to go: `grep -Eo pattern file | sort | uniq | wc`

Newdic

Newdic

- 20,000 words with various bits of information in tab-separated columns, e.g. transcription, stress, syllable count, spelling, frequency, part of speech.

Newdic

- 20,000 words with various bits of information in tab-separated columns, e.g. transcription, stress, syllable count, spelling, frequency, part of speech.
-

x	_	S1	a	23178	(N IA VB PP)
ardvark	'‘	S2	aardvark	0	(N)
xb@k	_'	S2	aback	2	(AV)
@bxkxs	'__	S3	abacus	0	(N)
xb@ft	_'	S2	abaft	0	(AV PP)
@bxloni	‘_'_	S4	abalone	0	(N)
xb@nd|n	_'_	S3	abandon	17	(VT N)
xbes	_'	S2	abase	0	(VT)
xb@S	_'	S2	abash	0	(VT)
xbet	_'	S2	abate	0	(VT VI)

How to count

How to count

- Easier because we can use the transcription in the first column.

How to count

- Easier because we can use the transcription in the first column.
- Easier because each word is on a separate line.

How to count

- Easier because we can use the transcription in the first column.
- Easier because each word is on a separate line.
- We need to count only those items where the relevant letter is at the beginning of the line.

Details

Details

- Mac or Unix
  - `grep -Ec '^[mn]' newdic.txt`
  - `grep -Ec '^[bd]' newdic.txt`

Details

- Mac or Unix
  - `grep -Ec '^[mn]' newdic.txt`
  - `grep -Ec '^[bd]' newdic.txt`
- Windows
  - `findstr "^[bd]" newdic.txt > bdwords.txt`
  - `findstr "^[mn]" newdic.txt > mnwords.txt`
  - Open resulting files with Notepad

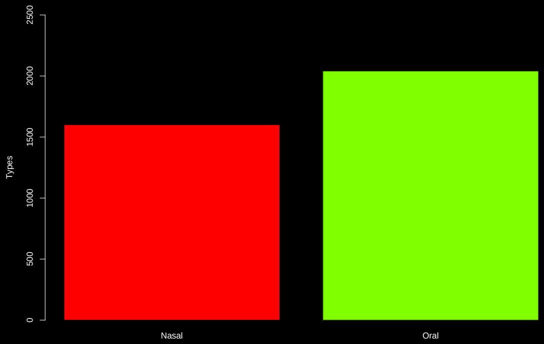
Results

Results

total number of words	19,528
words beginning with [bd]	2,041
words beginning with [mn]	1,600

Results

total number of words	19,528
words beginning with [bd]	2,041
words beginning with [mn]	1,600




```
> barplot(c(1600,2041))
```

Statistics

Statistics

-
```
> chisq.test(c(2041,1600))

        Chi-squared test for given probabilities

data:  c(2041, 1600)
X-squared = 53.4142, df = 1, p-value = 2.701e-13
```

Statistics

-
```
> chisq.test(c(2041,1600))

        Chi-squared test for given probabilities

data:  c(2041, 1600)
X-squared = 53.4142, df = 1, p-value = 2.701e-13
```
- There is a significant difference in the number of distinct words that begin with either a nasal or a voiced stop.

A third corpus example

A third corpus example

- Is there a difference in the number of types (distinct words) when we factor in the number of syllables in the word?

A third corpus example

- Is there a difference in the number of types (distinct words) when we factor in the number of syllables in the word?
- Again, we can do this better with a dictionary.

A third corpus example

- Is there a difference in the number of types (distinct words) when we factor in the number of syllables in the word?
- Again, we can do this better with a dictionary.
- The `newdic` dictionary is convenient as the third column indicates the number of syllables: S1, S2, S3, etc.

What to count

What to count

- We want to count the intersection of syllable size and initial consonant.

What to count

- We want to count the intersection of syllable size and initial consonant.
- For Unix or Mac: `grep -Ec '^[mn].*S1' newdic.txt`

What to count

- We want to count the intersection of syllable size and initial consonant.
- For Unix or Mac: `grep -Ec '^[mn].*S1' newdic.txt`
- For Windows: `findstr "^[mn].*S1" newdic.txt > mnwords.txt`, and open with Notepad.

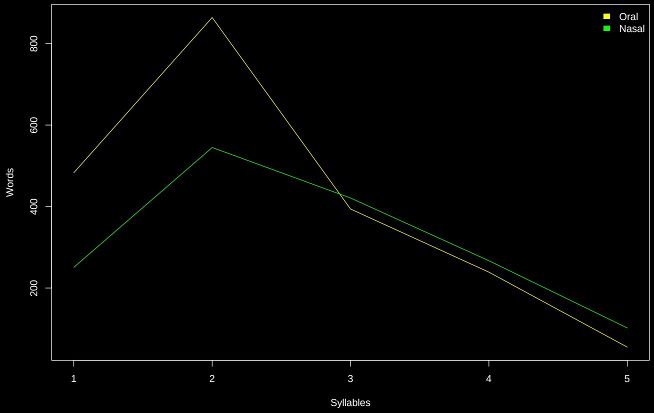
Results

Syllables	Nasal	Voiced
1	251	483
2	545	864
3	421	394
4	267	239
5	102	55

Results in a form R understands

Nasal	Voiced
251	483
545	864
421	394
267	239
102	55

Results graphed



Reading in and plotting the data

```
> sno <- read.table('sylnasor.txt',header=T)

> plot(sno$Voiced,type='l',col='yellow',
xlab='Syllables',ylab='Words')

> lines(sno$Nasal,col='green')

> legend('topright',legend=c('Oral','Nasal'),
fill=c('yellow','green'))
```

Stats

Stats

- One syllable:

```
> chisq.test(sno[1,])
X-squared = 73.3297, df = 1, p-value < 2.2e-16
```

Stats

- One syllable:

```
> chisq.test(sno[1,])
X-squared = 73.3297, df = 1, p-value < 2.2e-16
```

- Two syllables:

```
> chisq.test(sno[2,])
X-squared = 72.2221, df = 1, p-value < 2.2e-16
```

Stats

- One syllable:

```
> chisq.test(sno[1,])
X-squared = 73.3297, df = 1, p-value < 2.2e-16
```

- Two syllables:

```
> chisq.test(sno[2,])
X-squared = 72.2221, df = 1, p-value < 2.2e-16
```

- Three syllables:

```
> chisq.test(sno[3,])
X-squared = 0.8945, df = 1, p-value = 0.3443
```

Stats

- One syllable:

```
> chisq.test(sno[1,])
X-squared = 73.3297, df = 1, p-value < 2.2e-16
```

- Two syllables:

```
> chisq.test(sno[2,])
X-squared = 72.2221, df = 1, p-value < 2.2e-16
```

- Three syllables:

```
> chisq.test(sno[3,])
X-squared = 0.8945, df = 1, p-value = 0.3443
```

- Four syllables:

```
> chisq.test(sno[4,])
X-squared = 1.5494, df = 1, p-value = 0.2132
```

Stats

- One syllable:

```
> chisq.test(sno[1,])
X-squared = 73.3297, df = 1, p-value < 2.2e-16
```

- Two syllables:

```
> chisq.test(sno[2,])
X-squared = 72.2221, df = 1, p-value < 2.2e-16
```

- Three syllables:

```
> chisq.test(sno[3,])
X-squared = 0.8945, df = 1, p-value = 0.3443
```

- Four syllables:

```
> chisq.test(sno[4,])
X-squared = 1.5494, df = 1, p-value = 0.2132
```

- Five syllables:

```
> chisq.test(sno[5,])
X-squared = 14.0701, df = 1, p-value = 0.0001761
```

What does it all mean?

What does it all mean?

- There are significantly more instances of words that begin with voiced stops than that begin with nasals.

What does it all mean?

- There are significantly more instances of words that begin with voiced stops than that begin with nasals.
- There are significantly more distinct words that begin with voiced stops than that begin with nasals.

What does it all mean?

- There are significantly more instances of words that begin with voiced stops than that begin with nasals.
- There are significantly more distinct words that begin with voiced stops than that begin with nasals.
- The latter generalization holds for one- and two-syllable words, but does not hold for three- and four-syllable words. The pattern reverses itself with five-syllable words.

What does it all mean?

- There are significantly more instances of words that begin with voiced stops than that begin with nasals.
- There are significantly more distinct words that begin with voiced stops than that begin with nasals.
- The latter generalization holds for one- and two-syllable words, but does not hold for three- and four-syllable words. The pattern reverses itself with five-syllable words.
- Fame and fortune is yours if you can figure out why. . .

Doing experiments

Doing experiments

- Kinds of experiments

Doing experiments

- Kinds of experiments
- Materials

Doing experiments

- Kinds of experiments
- Materials
- Human subjects

Kinds of experiments

An experiment is a measure of human performance under some level of control.

Kinds of experiments

An experiment is a measure of human performance under some level of control.

- Pronounce words

Kinds of experiments

An experiment is a measure of human performance under some level of control.

- Pronounce words
- Collect judgments of grammaticality

Kinds of experiments

An experiment is a measure of human performance under some level of control.

- Pronounce words
- Collect judgments of grammaticality
- (Masked) priming

Kinds of experiments

An experiment is a measure of human performance under some level of control.

- Pronounce words
- Collect judgments of grammaticality
- (Masked) priming
- Speech perception

Kinds of experiments

An experiment is a measure of human performance under some level of control.

- Pronounce words
- Collect judgments of grammaticality
- (Masked) priming
- Speech perception
- Sentence processing

Kinds of experiments

An experiment is a measure of human performance under some level of control.

- Pronounce words
- Collect judgments of grammaticality
- (Masked) priming
- Speech perception
- Sentence processing

Are these relevant if one accepts a distinction between "competence" and "performance"?

Materials

Materials

- Select a factor: make a comparison, e.g. between words that begin with voiced oral stops and nasals.

Materials

- Select a factor: make a comparison, e.g. between words that begin with voiced oral stops and nasals.
- How many items in each group? Depends on the size of the effect, the amount of statistical noise expected, what statistics one will run, etc.

Materials

- Select a factor: make a comparison, e.g. between words that begin with voiced oral stops and nasals.
- How many items in each group? Depends on the size of the effect, the amount of statistical noise expected, what statistics one will run, etc.
- Sometimes one has to include irrelevant items: distractors. This is so subjects don't become aware of the contrast being investigated.

Human subjects

In the US, there is a huge hurdle at this point: IRB approval.

A hypothesis

Hx: people prefer words that begin with voiced stops to words that begin with nasals.

Experimental design

Experimental design

- Collect judgments on nonsense words individually and then compare values across sets

Experimental design

- Collect judgments on nonsense words individually and then compare values across sets
- Visual presentation, in English orthography

Experimental design

- Collect judgments on nonsense words individually and then compare values across sets
- Visual presentation, in English orthography
- Paired items

Experimental design

- Collect judgments on nonsense words individually and then compare values across sets
- Visual presentation, in English orthography
- Paired items
- Include distractors

Experimental items

	oral	nasal
labial	barna, bimp, bullow	marna, mimp, mullow
coronal	dift, dact, dilsey	nift, nact, nilsey

distractors: keff, farn, skelsat, starno, elso, plent

Free software

Free software

- DMDX (Windows)

Free software

- DMDX (Windows)
- Psyscope X (Mac)

Free software

- DMDX (Windows)
- Psyscope X (Mac)
- Web-based, e.g. html + javascript (+ php/sql), Webexp, etc.

Free software

- DMDX (Windows)
- Psyscope X (Mac)
- Web-based, e.g. html + javascript (+ php/sql), Webexp, etc.
- MiniJudge

Free software

- DMDX (Windows)
- Psyscope X (Mac)
- Web-based, e.g. html + javascript (+ php/sql), Webexp, etc.
- MiniJudge
- 3 $\times$ 5 cards

How I did it

How I did it

- Design and materials as above.

How I did it

- Design and materials as above.
- Put everything on a web page.

How I did it

- Design and materials as above.
- Put everything on a web page.
  - HTML to display items

How I did it

- Design and materials as above.
- Put everything on a web page.
  - HTML to display items
  - Javascript to randomize items and check responses

How I did it

- Design and materials as above.
- Put everything on a web page.
  - HTML to display items
  - Javascript to randomize items and check responses
  - Teeny perl program on the webserver emails results to me one by one.

How I did it

- Design and materials as above.
- Put everything on a web page.
  - HTML to display items
  - Javascript to randomize items and check responses
  - Teeny perl program on the webserver emails results to me one by one.

The version of the webpage that you can download is an offline version. It collects responses, but the experimenter must explicitly save them.

What we have to do

What we have to do

- Conceptually, we have to:
- Practically, we have to:

What we have to do

- Conceptually, we have to:
  - figure out what differences may have shown up; and
  - figure out if they are significant.
- Practically, we have to:
  - get the data into `R`;
  - calculate and plot the differences; and
  - figure out if they are significant.

Massaging the results

Massaging the results

- A form that we can get into R.

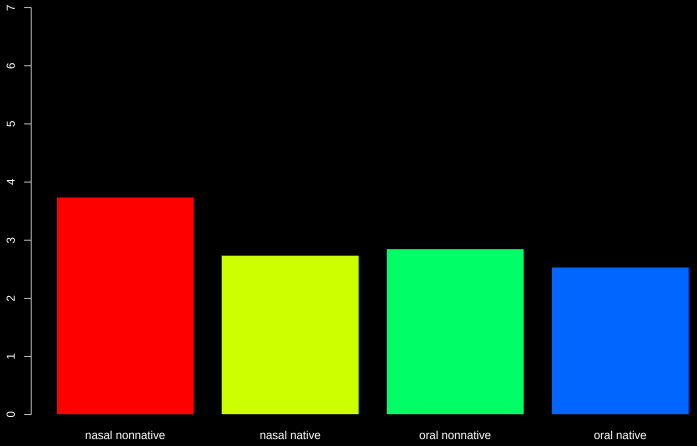
Massaging the results

- A form that we can get into R.
subject	native	place	nasal	item	response
s1	no	lab	nas	marna	4
s1	no	cor	nas	nact	1
s1	no	cor	nas	nilsey	2
s1	no	lab	oral	dact	1
s1	no	lab	oral	bimp	1
s1	no	cor	nas	nift	1
s1	no	lab	oral	barna	4
s1	no	lab	oral	dift	1
s1	no	lab	nas	mullow	1

Getting the data into R

```
> taiexp <- read.table('results.txt',header=T)
> summary(taiexp)
...
> attach(taiexp)
> tapply(response,list(native,nasal),mean)
         nas     oral
no  3.740741 2.851852
yes 2.739130 2.536232
```

Plotting the means



Creating that chart

```
> attach(taiexp)
> the.means <- tapply(response,list(native,nasal),mean)
> barplot(as.vector(the.means), ylim=c(0,7),col=rainbow(5),
names.arg=c('nasal nonnative', 'nasal native',
'oral nonnative','oral native'))
```

Statistics

Statistics

- There are differences, but are they meaningful?

Statistics

- There are differences, but are they meaningful?
- Here we must do an "Analysis of Variance" (ANOVA).

Statistics

- There are differences, but are they meaningful?
- Here we must do an “Analysis of Variance” (ANOVA).
- Warning: this is probably the wrong statistic to perform, but it’s what the literature uses, so it’s what we have to use.

ANOVA

ANOVA

- Determine all the variables and all the possible interactions of variables.

ANOVA

- Determine all the variables and all the possible interactions of variables.
- Factor out the random variables: subjects and items.

ANOVA

- Determine all the variables and all the possible interactions of variables.
- Factor out the random variables: subjects and items.
- Calculate the difference between conditions.

ANOVA

- Determine all the variables and all the possible interactions of variables.
- Factor out the random variables: subjects and items.
- Calculate the difference between conditions.
- Calculate the statistical noise (variance).

ANOVA

- Determine all the variables and all the possible interactions of variables.
- Factor out the random variables: subjects and items.
- Calculate the difference between conditions.
- Calculate the statistical noise (variance).
- Is the difference greater than the noise (F ratio)?

Factoring out subject noise

```
> summary(aov(response ~ native * nasal +
Error(subject/(native * nasal))))
Error: subject
          Df Sum Sq Mean Sq F value  Pr(>F)
native     1  39.18   39.18  5.1995 0.02664 *
Residuals 53 399.39    7.54
Error: subject:nasal
             Df Sum Sq Mean Sq F value   Pr(>F)
nasal         1 16.388  16.388 10.9964 0.001653 **
native:nasal  1 10.627  10.627  7.1306 0.010040 *
Residuals    53 78.986   1.490
```

Factoring out item noise

```
> summary(aov(response ~ native * nasal +
Error(item/(native * nasal))))
Error: item
          Df Sum Sq Mean Sq F value Pr(>F)
nasal      1  16.39   16.39  0.4965 0.4971
Residuals 10 330.08   33.01
Error: item:native
             Df Sum Sq Mean Sq F value   Pr(>F)
native        1 39.182  39.182 15.8293 0.002607 **
native:nasal  1 10.627  10.627  4.2931 0.065066 .
Residuals    10 24.753   2.475
```

A second experiment

Hx: people prefer words that begin with voiced stops to words that begin with nasals. This will show up with yes–no judgments.

How do we do this?

How do we do this?

- Web-based experient parallel to the last one, but changing the response categories to yes–no.

How do we do this?

- Web-based experient parallel to the last one, but changing the response categories to yes–no.
- Do we expect a different outcome?

How do we do this?

- Web-based experient parallel to the last one, but changing the response categories to yes–no.
- Do we expect a different outcome?
- 50 subjects participated.

Categorical response

Categorical response

- Yes–no judgments are a categorical response, not a quantitative response.

Categorical response

- Yes–no judgments are a categorical response, not a quantitative response.
- We need to convert our data into a quantitative measure so we can do an ANOVA.

Categorical response

- Yes–no judgments are a categorical response, not a quantitative response.
- We need to convert our data into a quantitative measure so we can do an ANOVA.
- Aggregate across places of articulation and multiple items within each condition.

Categorical response

- Yes–no judgments are a categorical response, not a quantitative response.
- We need to convert our data into a quantitative measure so we can do an ANOVA.
- Aggregate across places of articulation and multiple items within each condition.
- We can then do a by-subjects ANOVA.

Categorical response

- Yes–no judgments are a categorical response, not a quantitative response.
- We need to convert our data into a quantitative measure so we can do an ANOVA.
- Aggregate across places of articulation and multiple items within each condition.
- We can then do a by-subjects ANOVA.
- But we cannot do a by-items ANOVA.

Results massaged

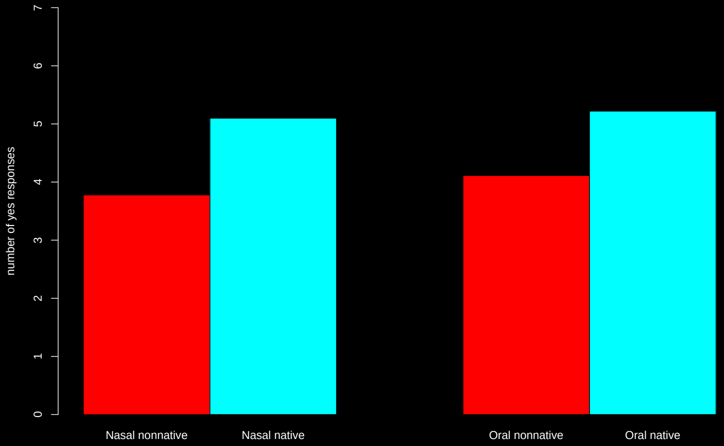
subject	native	nasal	response
s1	yes	nas	6
s1	yes	oral	6
s10	yes	nas	6
s10	yes	oral	6
s11	yes	nas	4
s11	yes	oral	4
s12	yes	nas	6
s12	yes	oral	6

Means

	nas	oral
no	3.777778	4.111111
yes	5.097561	5.219512

```
> te2 <- read.table('res2a.txt',header=T)
> attach(te2)
> tapply(response,list(native,nasal),mean)
```

Plotting the means



How we did that

```
> attach(te2)
> barplot(tapply(response,list(native,nasal),mean),
beside=T,ylim=c(0,7),col=rainbow(2),
names.arg=c('Nasal nonnative','Nasal native',
'Oral nonnative','Oral native'),
ylab='number of yes responses')
```

Stats

```
> summary(aov(response ~ native * nasal +
Error(subject/(native * nasal))))

Error: subject
          Df  Sum Sq Mean Sq F value Pr(>F)
native     1  21.757  21.757  7.6854 0.0079 **
Residuals 48 135.883   2.831
...
Error: subject:nasal
             Df  Sum Sq Mean Sq F value Pr(>F)
nasal         1  0.6400  0.6400  1.6004 0.2120
native:nasal  1  0.1649  0.1649  0.4123 0.5239
Residuals    48 19.1951  0.3999
```

Conclusions

Conclusions

- You need to figure out what you really want to do.

Conclusions

- You need to figure out what you really want to do.
- There are very important practical realities to doing any kind of experimental work, e.g. selecting a corpus, selecting the right tools, designing an experiment.

Conclusions

- You need to figure out what you really want to do.
- There are very important practical realities to doing any kind of experimental work, e.g. selecting a corpus, selecting the right tools, designing an experiment.
- You need some knowledge of statistics.

Conclusions

- You need to figure out what you really want to do.
- There are very important practical realities to doing any kind of experimental work, e.g. selecting a corpus, selecting the right tools, designing an experiment.
- You need some knowledge of statistics.
- We've barely scratched the surface here.