Discovering Words in Continuous Speech

Computational Simulations of Speech Segmentation

Çağrı Çöltekin

c.coltekin@rug.nl

Center for Language and Cognition
University of Groningen

June 2011

The puzzle to solve

ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj

The puzzle to solve

```
ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj
```

- No clear acoustic markers

The puzzle to solve

```
ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj
```

- No clear acoustic markers
- No lexical knowledge

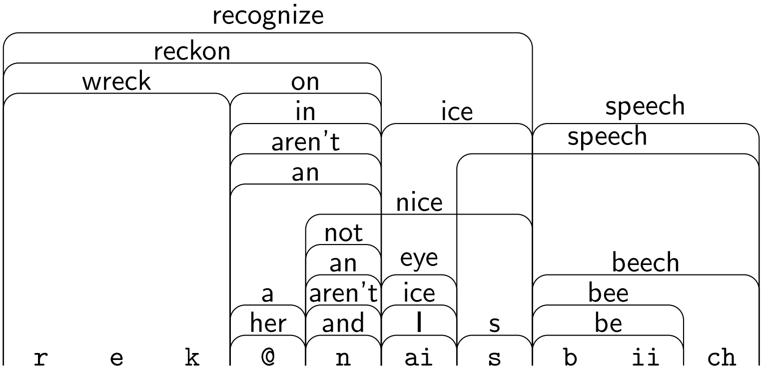
The puzzle to solve

```
ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj
```

- No clear acoustic markers
- No lexical knowledge
- Large acoustic variation
- Noise
- Even a comprehensive lexicon is enough

Recognize speech, or, wreck a nice beach

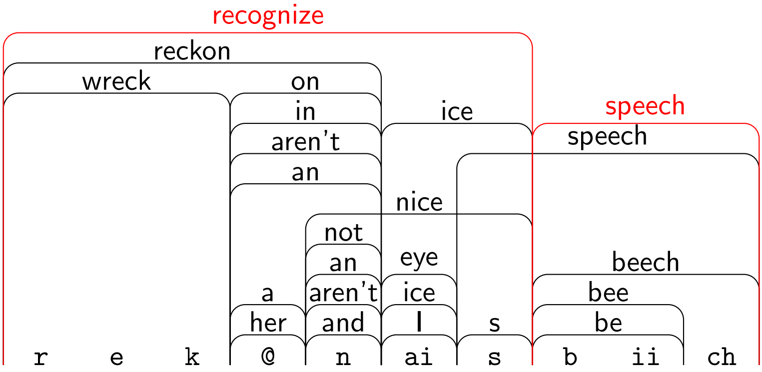
An automatic speech recognizers attempt to recognize the phrase ‘recognize speech’:



*Example reproduced from: (Shillcock, 1995)

Recognize speech, or, wreck a nice beach

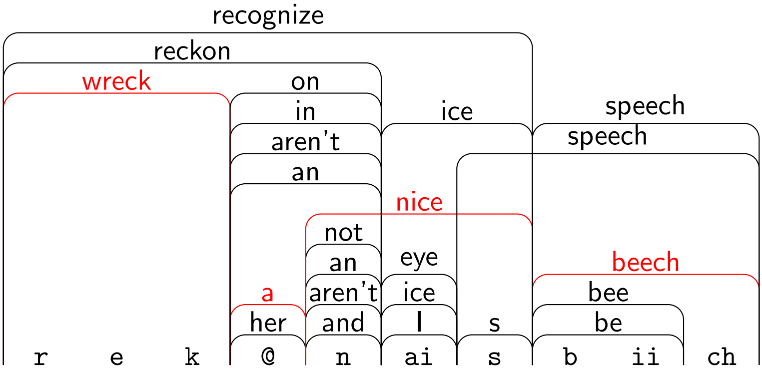
An automatic speech recognizers attempt to recognize the phrase 'recognize speech':



*Example reproduced from: (Shillcock, 1995)

Recognize speech, or, wreck a nice beach

An automatic speech recognizers attempt to recognize the phrase ‘recognize speech’:



*Example reproduced from: (Shillcock, 1995)

How do children segment?

A number of cues are well supporeted by psycholinguistic studies

- Acoustic cues
  - Lexical stress (Jusczyk et al., 1999)
  - Pauses (Wightman et al., 1992)
  - Allophonic alternations (Church, 1987).
  - Coarticulation (Johnson & Jusczyk, 2001).
  - Vowel/consonant harmony (Suomi et al., 1997; van Kampen et al., 2008).
- Words in isolation (Brent & Siskind, 2001)
- Phonotactics (Jusczyk et al., 1993)
- Statistical regularities (Saffran et al., 1996)

How do children segment?

A number of cues are well supporeted by psycholinguistic studies

- Acoustic cues
  - Lexical stress (Jusczyk et al., 1999)
  - Pauses (Wightman et al., 1992)
  - Allophonic alternations (Church, 1987).
  - Coarticulation (Johnson & Jusczyk, 2001).
  - Vowel/consonant harmony (Suomi et al., 1997; van Kampen et al., 2008).
- Words in isolation (Brent & Siskind, 2001)
- Phonotactics (Jusczyk et al., 1993)
- Statistical regularities (Saffran et al., 1996)

All these cues are useful, but none of them is enough by itself.

Statistical regularities for segmentation

Children very early in life (8-months) seem to be sensitive to statistical regularities between syllables (Saffran, Aslin, Newport 1996)

Statistical regularities for segmentation

Children very early in life (8-months) seem to be sensitive to statistical regularities between syllables (Saffran, Aslin, Newport 1996)

Training: `bidakupadotigolabubidakugolabupadoti...`

Statistical regularities for segmentation

Children very early in life (8-months) seem to be sensitive to statistical regularities between syllables (Saffran, Aslin, Newport 1996)

Training: `bidakupadotigolabubidakugolabupadoti...`

Statistical regularities for segmentation

Children very early in life (8-months) seem to be sensitive to statistical regularities between syllables (Saffran, Aslin, Newport 1996)

Training: bidakupadotigolabubidakugolabupadoti...

$TP(bi, da) = 1$ $TP(bu, pa) = \frac{1}{3}$

Statistical regularities for segmentation

Children very early in life (8-months) seem to be sensitive to statistical regularities between syllables (Saffran, Aslin, Newport 1996)

Training: `bidakupadotigolabubidakugolabupadoti...`

$TP(bi, da) = 1$

$TP(bu, pa) = \frac{1}{3}$

test G1: words

test G2: non-words

`padotibidakugolabupadoti ...`

`pagolabidotikugobdalaubu ...`

Statistical regularities for segmentation

Children very early in life (8-months) seem to be sensitive to statistical regularities between syllables (Saffran, Aslin, Newport 1996)

Training: bidakupadotigolabubidakugolabupadoti...

$TP(bi, da) = 1$ $TP(bu, pa) = \frac{1}{3}$

test G1: words

test G2: non-words

padotibidakugolabupadoti ...

pagolabidotikugobdalaubu ...

Children showed preference towards the 'words' that are used in the training phase.

back to the puzzle

```
ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj
```

back to the puzzle

```
ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj
```

Cues for the solution:

- Acoustic cues, such as *pauses*, *stress*, *coarticulation*, *allophonic alternations*, *vowel harmony*

back to the puzzle

ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj

Cues for the solution:

- Acoustic cues, such as *pauses*, *stress*, *coarticulation*, *allophonic alternations*, *vowel harmony*
- lexical knowledge

back to the puzzle

```
ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj
```

Cues for the solution:

- Acoustic cues, such as *pauses*, *stress*, *coarticulation*, *allophonic alternations*, *vowel harmony*
- lexical knowledge
- phonotactics

back to the puzzle

ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj

Cues for the solution:

- Acoustic cues, such as *pauses*, *stress*, *coarticulation*, *allophonic alternations*, *vowel harmony*
- lexical knowledge
- phonotactics
- utterance boundaries

back to the puzzle

ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj

Cues for the solution:

- Acoustic cues, such as *pauses*, *stress*, *coarticulation*, *allophonic alternations*, *vowel harmony*
- lexical knowledge
- phonotactics
- utterance boundaries
- distributional regularities

back to the puzzle

```
ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj
```

Cues for the solution:

- Acoustic cues, such as *pauses*, *stress*, *coarticulation*, *allophonic alternations*, *vowel harmony*
- lexical knowledge
- phonotactics
- utterance boundaries
- distributional regularities
- predictability

back to the puzzle

```
ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeph
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczcjsejf
cbczcjsejf
zpvepoumjlfuibupof
plbznpnnzublfuijtpvu
dpx
uifdpxtbztnppnpp
xibuepftuifdpxtbzopnj
```

Cues for the solution:

- Acoustic cues, such as *pauses*, *stress*, *coarticulation*, *allophonic alternations*, *vowel harmony*
- lexical knowledge
- phonotactics
- utterance boundaries
- distributional regularities
- predictability
  $\mathrm{TP}(\mathrm{ju}) = 11/27$
  $\mathrm{TP}(\mathrm{zu}) = 2/23$

Predictability

Predictability within units is high, predictability between units is low.

Given a sequence `lr`, where `l` and `r` are sequences of phonemes:

- If `l` help us predict `r`, `lr` is likely to be part of a word.
- If observing `r` after `l` is surprising it is likley that there is a boundary between `l` and `r`.

Predictability

Predictability within units is high, predictability between units is low.

Given a sequence `lr`, where `l` and `r` are sequences of phonemes:

- If `l` help us predict `r`, `lr` is likely to be part of a word.
- If observing `r` after `l` is surprising it is likley that there is a boundary between `l` and `r`.

An old method based on predictability, called *successor variety* (SV), is suggested by Haris (1955):

The morpheme boundaries are at the locations where there is a high variety of possible phonemes that follow the initial segment.

Predictability

Predictability within units is high, predictability between units is low.

Given a sequence `lr`, where `l` and `r` are sequences of phonemes:

- If `l` help us predict `r`, `lr` is likely to be part of a word.
- If observing `r` after `l` is surprising it is likley that there is a boundary between `l` and `r`.

An old method based on predictability, called *successor variety* (SV), is suggested by Haris (1955):

The morpheme boundaries are at the locations where there is a high variety of possible phonemes that follow the initial segment.

read- readi-

Predictability

Predictability within units is high, predictability between units is low.

Given a sequence `lr`, where `l` and `r` are sequences of phonemes:

- If `l` help us predict `r`, `lr` is likely to be part of a word.
- If observing `r` after `l` is surprising it is likley that there is a boundary between `l` and `r`.

An old method based on predictability, called *successor variety* (SV), is suggested by Haris (1955):

The morpheme boundaries are at the locations where there is a high variety of possible phonemes that follow the initial segment.

read-{a,e,i,j,o,s,y,-,',$} (10)
read
reads
reading
readjusted
...

readi-{e, l, n} (3)
reading
readied
readily

Measures of (un)predictability

- Transitional probability

$$\mathrm{TP}(\mathtt{l},\mathtt{r}) = \frac{P(\mathtt{lr})}{P(\mathtt{l})}$$

- Pointwise mutual information

$$\mathrm{MI}(\mathtt{l},\mathtt{r}) = log_2 \frac{P(\mathtt{lr})}{P(\mathtt{l})P(\mathtt{r})}$$

- Successor value

$$\mathrm{SV}(\mathtt{l}) = \sum_{\mathtt{r}\in A} c(\mathtt{l},\mathtt{r})$$

- Boundary entropy

$$\mathrm{H}(\mathtt{l}) = -\sum_{\mathtt{r}\in A} P(\mathtt{r}|\mathtt{l})\ log_2 P\left(\mathtt{r}|\mathtt{l}\right)$$

All these measures are related, but they are not the same.

This list is not exclusive, there are other measures of (un)predictability.

Two alternations

- The assymmetric measures, TP, SV, H, have their ‘reverse’ counterparts.
  - It may sound odd to predict the past (`l`), using the future (`r`), people seem to make use of this type or dependencies.
- The size of `l` and `r` in calculations matter.
  - The `l` and `r` we use in calculation can be phonme sequences of arbitrary size.
  - Sequences of different sizes capture different linguistic generalizations.

How to Calculate the Measures

`#   I   z   D   &   t   6   k   I   t   i   #`

How to Calculate the Measures

`#  I | z  D  &  t  6  k  I  t  i  #`

0.40

$$TP(\texttt{\#I}, \texttt{z}) = P(\texttt{z}|\texttt{\#I}) = 0.40$$

How to Calculate the Measures

| I z | D | & t 6 k I t i

0.40 0.22

$$TP(\mathtt{Iz}, \mathtt{D}) = P(\mathtt{D}|\mathtt{Iz}) = 0.22$$

How to Calculate the Measures

I z D & t 6 k I t i

0.40 0.22 0.46

$$TP(\mathtt{zD}, \mathtt{\&}) = P(\mathtt{\&}|\mathtt{zD}) = 0.46$$

How to Calculate the Measures

#	I	z	D	&	t	6	k	I	t	i	#

0.40 0.22 0.46 0.99

$$TP(\texttt{D\&}, \texttt{t}) = P(\texttt{t}|\texttt{D\&}) = 0.99$$

How to Calculate the Measures

#	I	z	D	&	t	6	k	I	t	i	#
	0.40	0.22	0.46	0.99	0.03						

$$TP(\texttt{\&t}, 6) = P(6|\texttt{\&t}) = 0.03$$

How to Calculate the Measures

`#  I  z  D  &  t  6  k  I  t  i  #`

0.40 0.22 0.46 0.99 0.03 0.04

$$TP(\texttt{t6}, \texttt{k}) = P(\texttt{k}|\texttt{t6}) = 0.04$$

How to Calculate the Measures

$$\begin{array}{ccccccccccccccccccccccccc}
\# & & \texttt{I} & & \texttt{z} & & \texttt{D} & & \texttt{\&} & & \texttt{t} & & \boxed{\texttt{6} \quad \texttt{k}} & & \boxed{\texttt{I}} & & \texttt{t} & & \texttt{i} & & \# \\
& 0.40 & & 0.22 & & 0.46 & & 0.99 & & 0.03 & & 0.04 & & 0.30 & & & & & & &
\end{array}$$

$$TP(\texttt{6k}, \texttt{I}) = P(\texttt{I}|\texttt{6k}) = 0.30$$

How to Calculate the Measures

`#` `I` `z` `D` `&` `t` `6` `k` `I` `t` `i` `#`

0.40 0.22 0.46 0.99 0.03 0.04 0.30 0.48

$$TP(\texttt{kI}, \texttt{t}) = P(\texttt{t}|\texttt{kI}) = 0.48$$

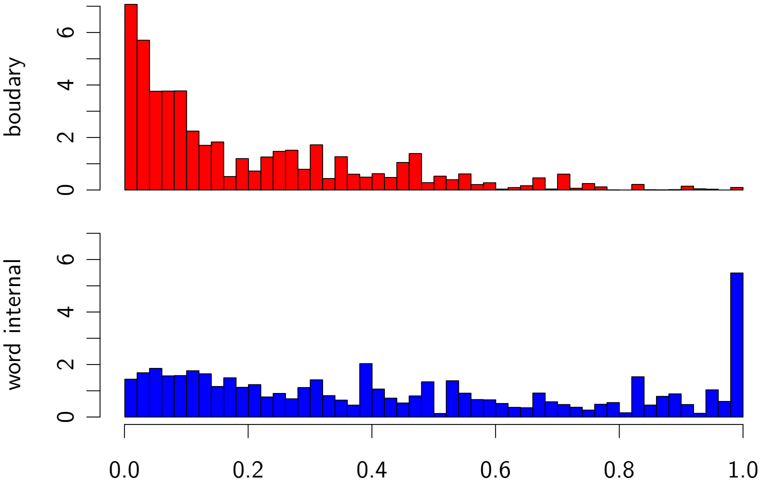
How to Calculate the Measures

#	I	z	D	&	t	6	k	I	t	i	#
	0.40	0.22	0.46	0.99	0.03	0.04	0.30	0.48	0.10		

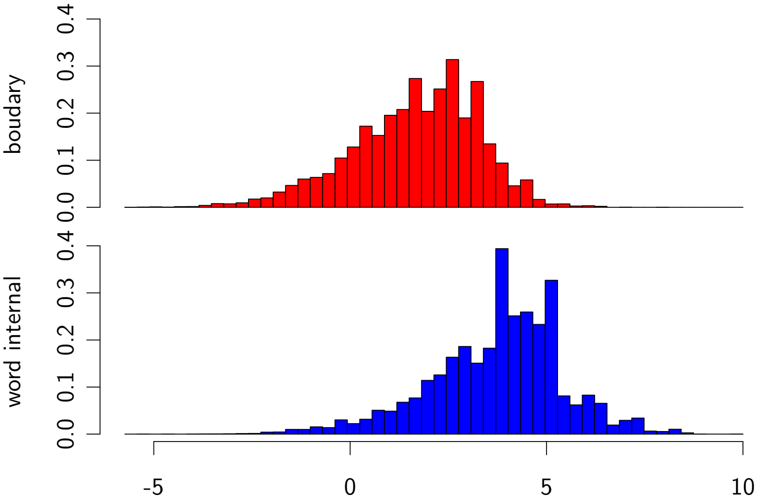
$$TP(\mathtt{It}, \mathtt{i}) = P(\mathtt{i}|\mathtt{It}) = 0.10$$

Calculations are done on a corpus of child-directed English

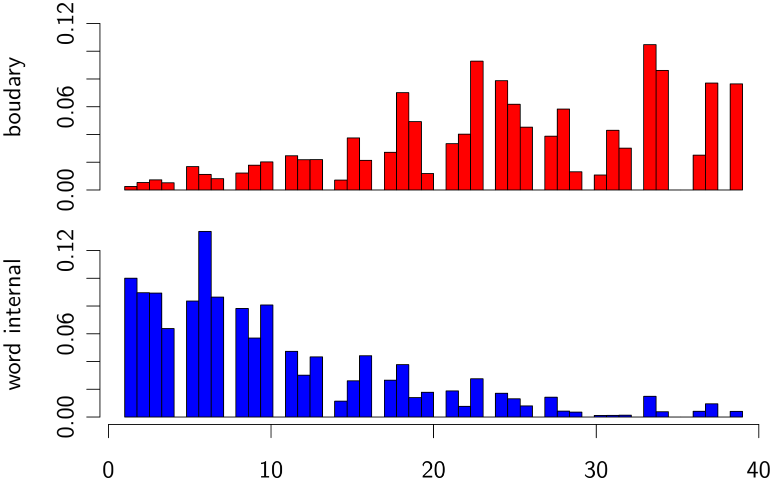
Transitional Probability



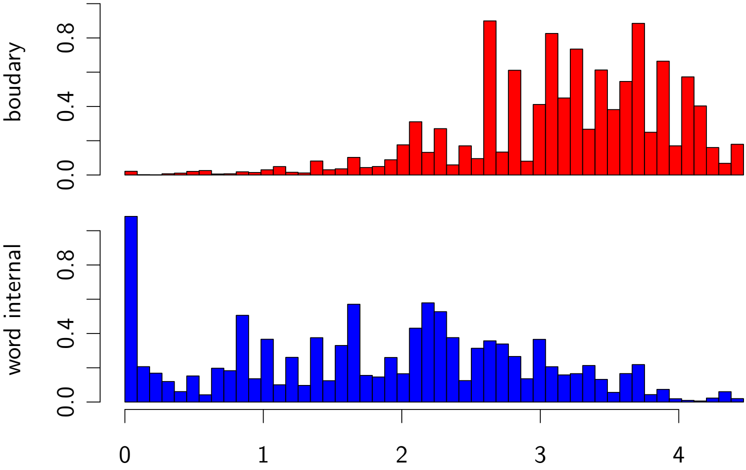
Pointwise Mutual Information



Successor Variety



Entropy



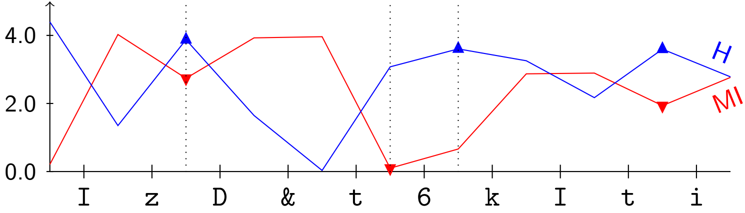
An unsupervised method

- An obvious way to segment the sequence is using a threshold value. However, the choice of threshold is difficult in an unsupervised system.

An unsupervised method

- An obvious way to segment the sequence is using a threshold value. However, the choice of threshold is difficult in an unsupervised system.

A simple unsupervised method: segment at peaks/valleys.



Combining multiple measures: a simple method

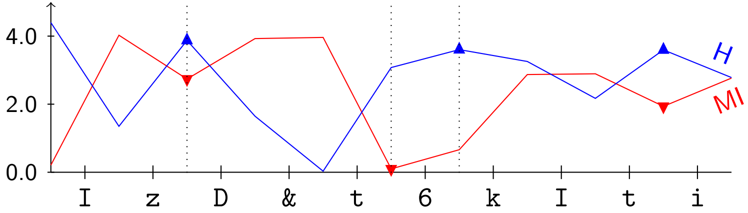
Majority voting (algorithm) works if

1. Votes are cast (relatively) independently.
2. Decisions of the voters are better than random.

Combining multiple measures: a simple method

Majority voting (algorithm) works if

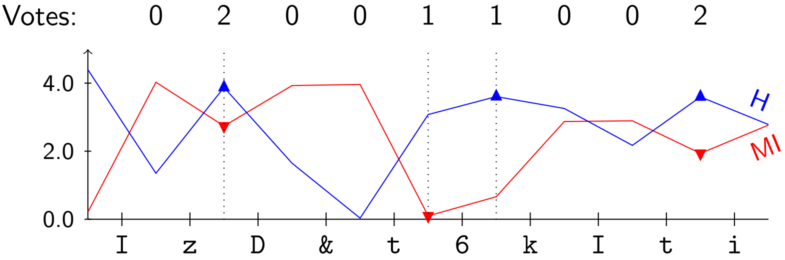
1. Votes are cast (relatively) independently.
2. Decisions of the voters are better than random.



Combining multiple measures: a simple method

Majority voting (algorithm) works if

1. Votes are cast (relatively) independently.
2. Decisions of the voters are better than random.



Evaluation

We evaluate the results agianst a linguistically motivated 'gold standard'.

- True positives (TP), or hits, are the items that match with the gold standard.
- False positives (FP), or fals alarms, are the items that the procedure found, but do not exist in the gold standard.
- False negatives (FN), or misses, are the items that are in the gold standard that our procedure did not find.

$$\text{precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \qquad \text{recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\text{f-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Boundary, type or token scores

In case of segmentation 'items' to count can be either *boundaries*, word *tokens* or word *types*.

An example:

Input	`kittysayitagainlovekitty`
Gold standard	`kitty say it again love kitty`
Segmentation	`kitty sayit aga inlove kitty`

	TP	FP	FN	precision	recall	f-score
Boundary (B)	3	1	2	$3/4 = 0.75$	$3/5 = 0.60$	0.66
Token (W)	2	3	3	$2/5 = 0.40$	$2/6 = 0.33$	0.35
Type (L)	1	3	3	$1/4 = 0.25$	$1/4 = 0.20$	0.22

So far...

- We have a set of measures: TP, MI, SV, H, and reverse versions of TP, SV, H.
- These measures are similar, yet do not always indicate the same result.
- All of the measures can be computed using a varying window size over the input sequence.
- A certain measure calculated with different window sizes are expected to capture different linguistic generalizations.
- We have an unsupervised procedure to guess boundaries: peak (or valley) strategy.
- We have a simple method to combine multiple decisions: majority voting.
- We also have an evaluation scheme, given a gold standard, we can measure our success using precision and recall.

Putting all together

Algorithm 1: A simple algorithm using multiple measures

foreach *utterance* **do**
    foreach *phoneme position in the utterance* **do**
        Get the majority vote of all measures calculated using context sizes one to four;
        if *majority vote is positive* **then**
            insert a boundary;
    output the segmented utterance ;

The results

method	BP	BR	BF	WP	WR	WF	LP	LR	LF
Random	27.4	27.4	27.4	12.7	12.7	12.7	6.4	46.4	11.3
Predictability	92.7	76.0	83.5	77.2	67.4	72.0	28.4	65.1	39.5
Baseline	84.2	82.7	83.5	72.1	71.2	71.6	50.7	61.1	55.4

Results are obtained using Algorithm 1 on phonemic transcriptions of child directed speech from Berstain-Ratner corpus.

Segmentation puzzle: a solution

```
ljuuz uibut sjhiu ljuuz
ljuuz tbz ju bhbjo mpwf ljuuz
xibut uibu
ljuuz
ep zpv xbou npsf njml ipofz
ljuuz ljuuz ephhjf
opnj xibu epft b ljuuz tbz
xibu epft b ljuuz tbz
ephhjf eph
ephhjf
opnj xibu epft uif ephhjf tbz
xibu epft uif ephhjf tbz
mjuumf cbcz cjsejf
cbcz cjsejf
zpv epou mjlf uibu pof
plbz npnnz ublf uijt pvu
dpx
uif dpx tbzt npp npp
xibu epft uif dpx tbz opnj
```

Segmentation puzzle: a solution

ljuuz uibut sjhiu ljuuz	ljuuz uibu tsjhiuljuuz
ljuuz tbz ju bhbjo mpwf ljuuz	ljuuz tbz jubhbjompwfljuuz
xibut uibu	xibu tuibu
ljuuz	ljuuz
ep zpv xbou npsf njml ipofz	ep zpvxbounpsfnjmli pof z
ljuuz ljuuz ephhjf	ljuuz ljuuz ephhjf
opnj xibu epft b ljuuz tbz	opnj xibu ep ftb ljuuz tbz
xibu epft b ljuuz tbz	xibu ep ftb ljuuz tbz
ephhjf eph	ephhjf eph
ephhjf	ephhjf
opnj xibu epft uif ephhjf tbz	opnj xibu epft uif ephhjf tbz
xibu epft uif ephhjf tbz	xibu ep ft uif ephhjf tbz
mjuumf cbcz cjsejf	mjuumfcbczcjsejf
cbcz cjsejf	cbczcjsejf
zpv epou mjlf uibu pof	zpv epoumj lf uibu pof
plbz npnnz ublf uijt pvu	plbznpnnzublfui jtpvu
dpx	dpx
uif dpx tbzt npp npp	uif dpx tbz tnppnpp
xibu epft uif dpx tbz opnj	xibu epft uif dpx tbz opnj

Segmentation puzzle: a solution

kitty thats right kitty	kitty that srightkitty
kitty say it again love kitty	kitty say itagainlovekitty
whats that	what sthat
kitty	kitty
do you want more milk honey	do youwantmoremilkh one y
kitty kitty doggie	kitty kitty doggie
nomi what does a kitty say	nomi what do esa kitty say
what does a kitty say	what do esa kitty say
doggie dog	doggie dog
doggie	doggie
nomi what does the doggie say	nomi what does the doggie say
what does the doggie say	what do es the doggie say
little baby birdie	littlebabybirdie
baby birdie	babybirdie
you dont like that one	you dontli ke that one
okay mommy take this out	okaymommytaketh isout
cow	cow
the cow says moo moo	the cow say smoomoo
what does the cow say nomi	what does the cow say nomi

Summary

The segmentation procedure we have just reviewed

- is in line with the psycholinguistic research,
- is completely unsupervised,
- is incremental,
- performs competitive with an alternative state of the art segmentation method.

This is only the part of the solution, we can

- use information from utterance boundaries,
- keep an explicit lexicon and use it for further segmentation,
- make use of acoustic cues,
- use a better algorithm for boundary guessing.

Appendix

Computational Models of Language Acquisition

Answers to the questions on language acquisition should eventually come from neuroscience. But we seem to be far from this yet.

- Computational models can be used to test the validity of theories under various conditions.
- Computational models require theories to be described explicitly.
- Compared to other formal aproaches, computational simulations can benefit from using real data (e.g. CHILDES database).

Computational Models for Language Acquisition

Computational models of human language acquisition has to meet some criteria that is not always applicable for engineering oriented CL applications.

- Modes should use realistic input, such as naturally occurring child directed speech.
- Any additional source of information, or heuristics should be justifiable.
- Learning should proceed on-line: models should not require all the input data available at once.
- Models should not pose unrealistic bounds on memory and computation resources.
- The assumptions and predictions of the model should match (at least should not conflict with) psycholinguistic evidence.

Reverse measures

The assymmetric measures also have ‘backwards’ counterparts:

- Reverse transitional probability

$$\mathrm{TP}_r(\mathtt{l},\mathtt{r}) = P(\mathtt{l}|\mathtt{r}) = \frac{P(\mathtt{lr})}{P(\mathtt{r})} \approx \frac{\mathit{frequency}(\mathtt{lr})}{\mathit{frequency}(\mathtt{r})} \qquad (1)$$

- Reverse successor variety (predecessor variety)

$$\mathrm{SV}_r(\mathtt{r}) = \sum_{\mathtt{l}\in A} c(\mathtt{l},\mathtt{r}) \qquad (2)$$

  where,

$$c(\mathtt{l},\mathtt{r}) = \begin{cases} 1 & \text{if substring } \mathtt{lr} \text{ occurs in the corpus} \\ 0 & \text{otherwise} \end{cases}$$

  and A is the set of phonemes (the alphabet).

- Reverse boundary entropy

$$\mathrm{H}_r(\mathtt{r}) = -\sum_{\mathtt{l}\in A} P(\mathtt{l}|\mathtt{r})\, \mathit{log}_2 P(\mathtt{l}|\mathtt{r}) \qquad (3)$$

Difficulties of learning segmentation

- No clear acouistic markers in fluent speech.
- Large speaker variation in acoustic input.
- Noise in the environmet.
- Children has to start with no knwledge of words.
- Even with a comprehensive knowledge of words, segmentation is still difficult because of multiple plausible segmentations.
  For example:
  `/6go/`: `/6go/` 'ago' or `/6 go/` 'a go'?
  `/Itsnoz/`: `/Its noz/` 'its nose' or `/It snoz/` 'it snows'?