EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

PHILOSOPHISCHE
FAKULTÄT
Seminar für Sprachwissenschaft

Tackling a grand challenge in the visualization of language and linguistic data

Chris Culy

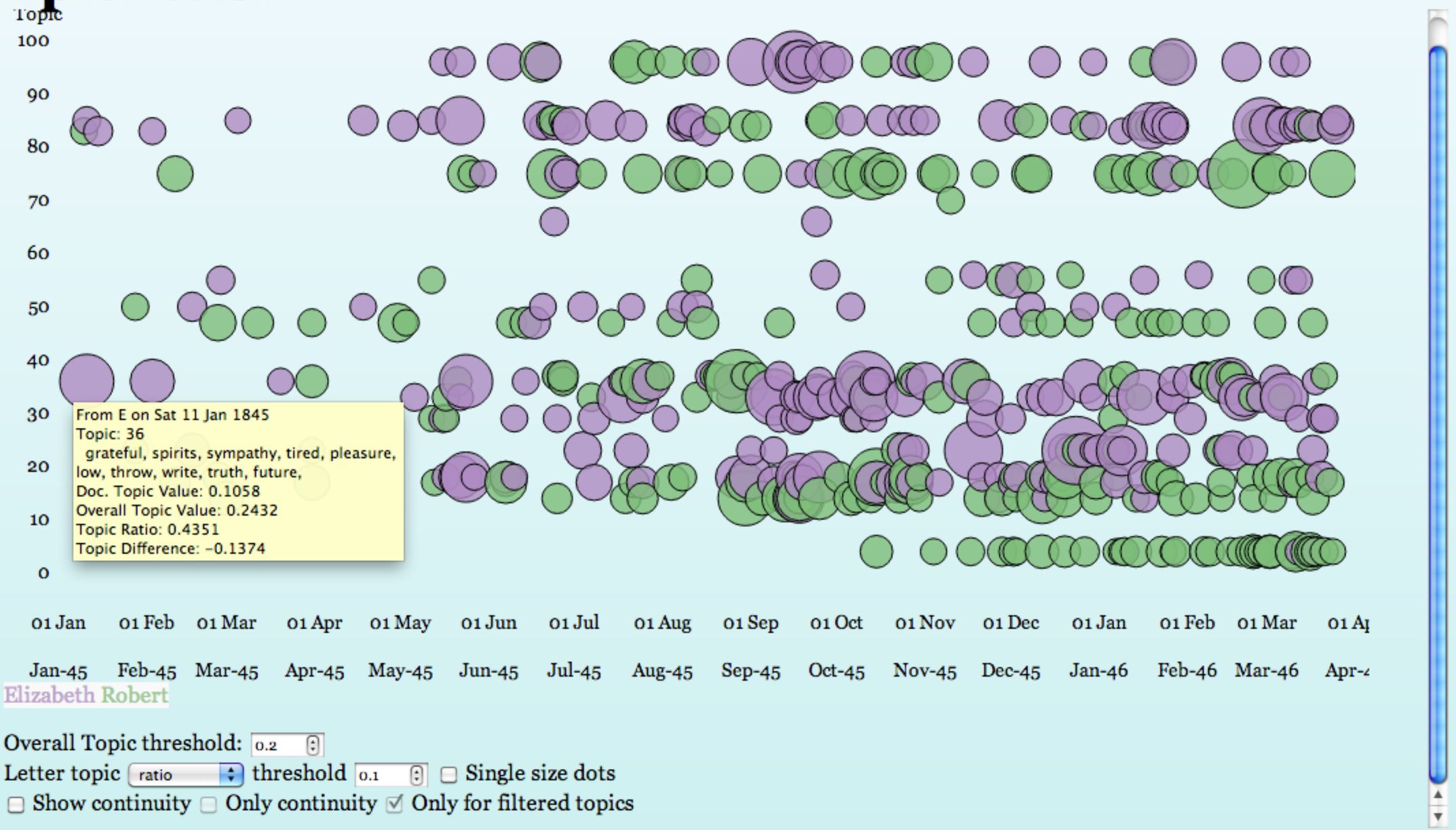
DGfS 2013 Workshop on the Visualization of Linguistic Patterns

13 March 2013

Visualizations put ideas into our heads ...

3 Separate goals of visualizations

1. Exploration

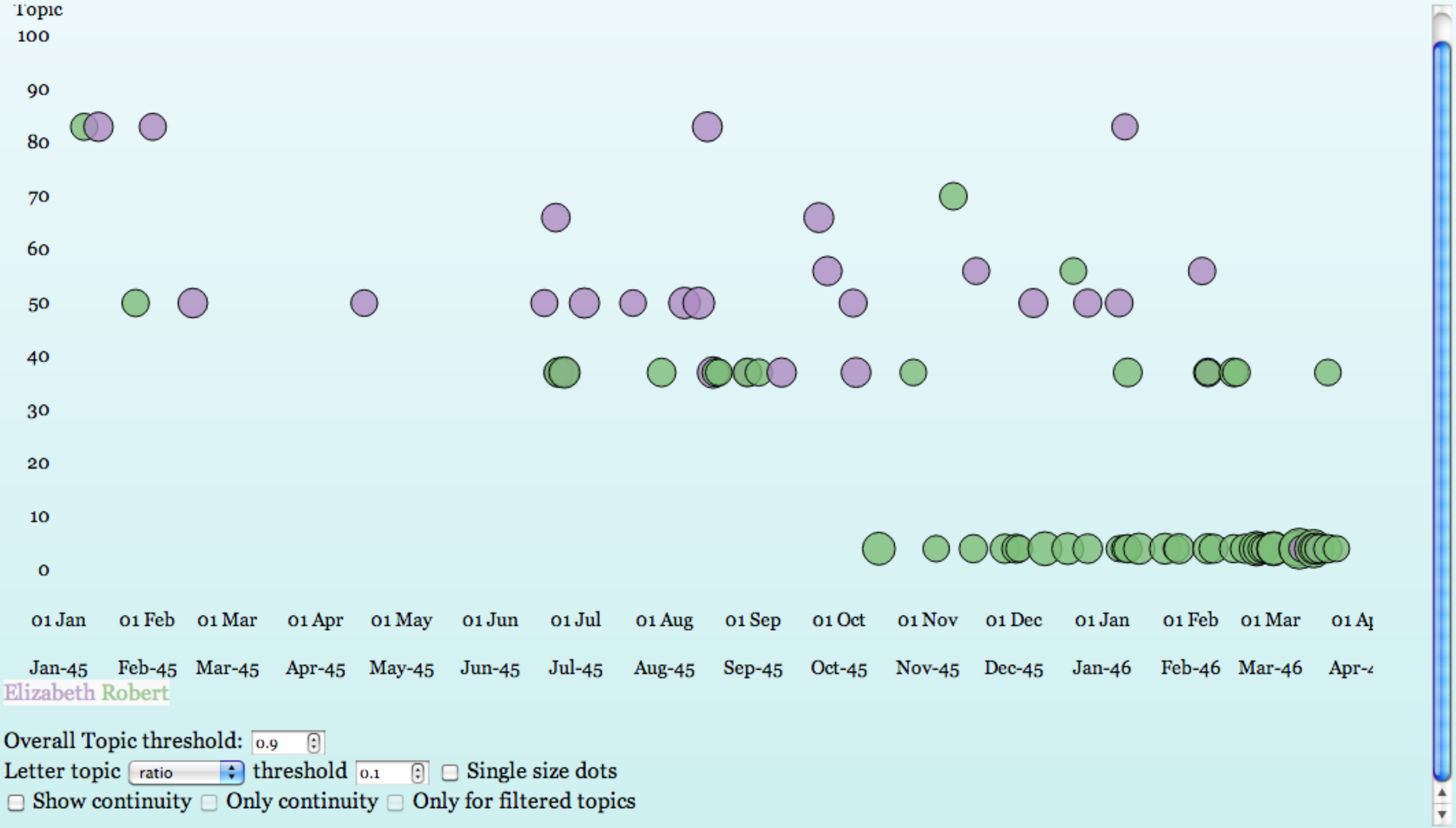


3 Separate goals of visualizations

2. Comprehension (Analysis)

3 Separate goals of visualizations

3. Communication (Presentation)



3 Separate goals of visualizations

1. Exploration
2. Comprehension (Analysis)
3. Communication (Presentation)

3 Separate goals of visualizations

1. Exploration
2. Comprehension (Analysis)
3. Communication (Presentation)

Research Question: Which visualizations are best suited for which of these goals?

What do users want?

1. **Grinstein's Grand Challenge**

Sources: Georges Grinstein, Tableau, Spotfire

What do users want?

1. Grinstein's Grand Challenge
2. **Pre-existing first steps**
    - Tableau
    - Spotfire

Sources: Georges Grinstein, Tableau, Spotfire

What do users want?

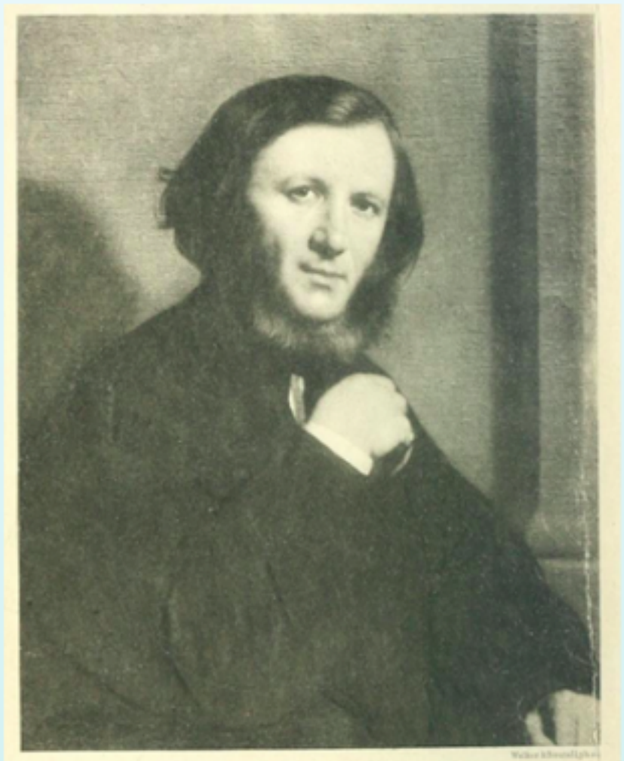
1. **Grinstein's Grand Challenge**
2. **Pre-existing first steps**
    - Tableau
    - Spotfire
3. **Limitations**
    - Simple data: numbers, dates, geographic, categories
    - No notion of task
    - No notion of preferences

Sources: Georges Grinstein, Tableau, Spotfire

Tackling the Grand Challenge

Dataset genres

Person-oriented correspondence



Robert Browning

Elizabeth Barrett Browning

Sources: on archive.org, on archive.org

Tackling the Grand Challenge

Dataset genres

Person-oriented correspondence

Robert Browning

Elizabeth Barrett Browning

What other dataset genres are there?

Sources: on archive.org, on archive.org

What makes L/L data different?

1. **Language is not *mappable***

What makes L/L data different?

1. Language is not *mappable*
2. **Strings are special**

What makes L/L data different?

1. Language is not *mappable*
2. Strings are special
3. **Individual pieces of L/L data are meaningful**

What makes L/L data different?

1. Language is not *mappable*
2. Strings are special
3. Individual pieces of L/L data are meaningful
4. **Much *linguistic* data is computed, not observed**

Strings are different
Data is meaningful

Num	Date	From	-5	-4	-3	-2	-1	hit	+1	+2	+3	+4	+5
1	1845-01-10	R	Hatcham/NP	,/,	Surrey/NP	./SENT	I/PP	**love/VBP**	your/PP$	verses/NNS	with/IN	all/PDT	my/PP$
1	1845-01-10	R	,/,	as/IN	I/PP	say/VBP	,/,	**love/VBP**	these/DT	books/NNS	with/IN	all/PDT	my/PP$
1	1845-01-10	R	my/PP$	heart/NN	--/:	and/CC	I/PP	**love/VBP**	you/PP	too/RB	./SENT	Do/VBP	you/PP
5	1845-01-28	R	and/CC	dissertate/VB	upon/RP	that/IN	I/PP	**love/VBP**	most/JJS	and/CC	least/JJS	--/:	I/PP
7	1845-02-11	R	so/RB	,/,	if/IN	"/``	I/PP	**love/VBP**	you/PP	"/``	were/VBD	always/RB	outspoken/JJ
8	1845-02-17	E	not/RB	see/VB	where/WRB	./SENT	I/PP	**love/VBP**	the/DT	drama/NN	too/RB	./SENT	I/PP
8	1845-02-17	E	princes/NNS	in/IN	poetry/NN	./SENT	I/PP	**love/VBP**	them/PP	through/IN	all/PDT	the/DT	deeps/NNS
9	1845-02-26	R	"/"	so/RB	thoroughly/RB	does/VBZ	he/PP	**love/VBP**	and/CC	live/VBP	by/IN	it/PP	./SENT
13	1845-03-12	R	properly/RB	for/IN	it/PP	,/,	I/PP	**love/VBP**	and/CC	wish/VBP	you/PP	well/RB	!/SENT
17	1845-04-18	E	you/PP	--/:	and/CC	whatever/WDT	you/PP	**love/VBP**	or/CC	hate/VBP	,/,	whatever/WDT	charms/VBZ
19	1845-05-02	E	Mr./NP	Browning/NP	,/,	that/WDT	we/PP	**love/VBP**	the/DT	darkness/NN	and/CC	use/VB	a/DT
30	1845-05-24	E	of/IN	my/PP$	aunts/NNS	whom/WP	I/PP	**love/VBP**	,/,	and/CC	have/VBP	not/RB	met/VBN
32	1845-05-25	E	every/DT	word/NN	./SENT	"/``	I/PP	**love/VBP**	the/DT	truth/NN	and/CC	can/MD	bear/VB
59	1845-07-09	R	And/CC	I/PP	,/,	too/RB	,/,	**love/VBP**	to/TO	have/VB	few/JJ	friends/NNS	,/,
87	1845-08-25	E	I/PP	,/,	for/IN	one/CD	,/,	**love/VBP**	him/PP	!/SENT	and/CC	when/WRB	,/,
102	1845-09-13	R	as/IN	I/PP	observed/VBD	,/,	I/PP	**love/VBP**	you/PP	as/IN	you/PP	now/RB	are/VBP
126	1845-10-13	R	bless/VBP	you/PP	and/CC	all/RB	you/PP	**love/VBP**	!/SENT	dearest/JJS	,/,	I/PP	am/VBP
128	1845-10-15	R	,/,	my/PP$	own/JJ	,/,	dearest/JJS	**love/VBP**	,/,	that/IN	this/DT	is/VBZ	for/IN
136	1845-10-23	R	in/IN	the/DT	least/JJS	./SENT	I/PP	**love/VBP**	you/PP	because/IN	I/PP	love/VBP	you/PP
136	1845-10-23	R	I/PP	love/VBP	you/PP	because/IN	I/PP	**love/VBP**	you/PP	;/:	I/PP	see/VBP	you/PP
136	1845-10-23	R	,/,	live/VBP	my/PP$	life/NN	,/,	**love/VBP**	my/PP$	love/NN	./SENT	When/WRB	I/PP
138	1845-10-25	E	loved/VBD	him/PP	tenderly/RB	(/(	and/CC	**love/VBP**	him/PP	)/)	,/,	.../:	and/CC
144	1845-11-04	R	love/NN	;/:	not/RB	as/IN	I/PP	**love/VBP**	you/PP	--/:	not/RB	for/IN	--/:
166	1845-12-04	E	if/IN	I/PP	feel/VBP	that/IN	you/PP	**love/VBP**	me/PP	,/,	can/MD	I/PP	help/VB
169	1845-12-08	E	ripen/VB	the/DT	knowledge/NN	./SENT	They/PP	**love/VBP**	Tennyson/NP	so/RB	much/RB	that/IN	the/DT
177	1845-12-19	R	to/TO	avoid/VB	writing/VBG	that/IN	I/PP	**love/VBP**	and/CC	love/VBP	and/CC	love/VBP	again/RB

Strings are different
Data is meaningful

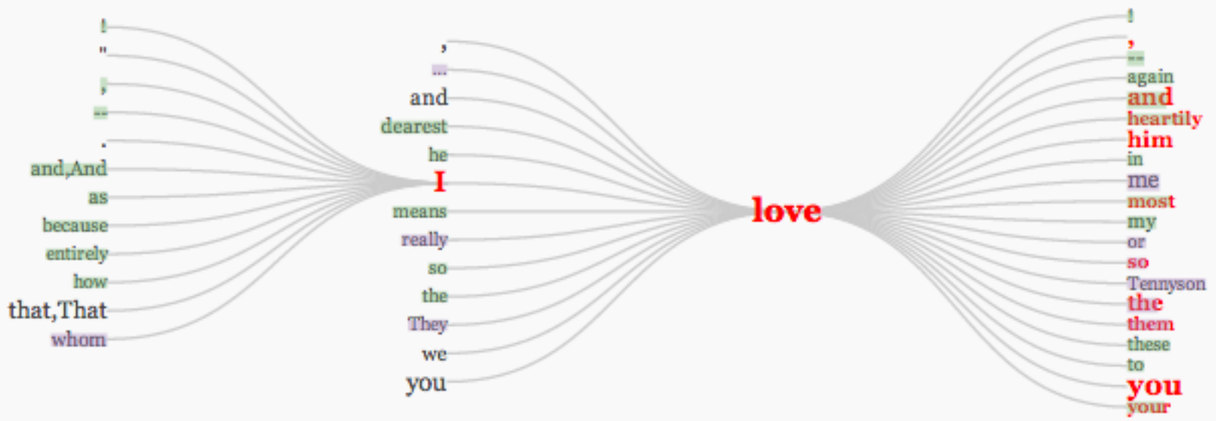
Input file: Choose File No file chosen Word to use: love/VBP

Sort branches by: token POS branching

Limit first left POS to:

Limit first right POS to: Update filters

Find in tree: [aA] Clear Show POS



Uses: DoubleTreeJS

Research questions about data

What are the types of data that are especially relevant to L/L?

Research questions about data

What are the types of data that are especially relevant to L/L?

What specialized visualizations are especially relevant to L/L data?

Research questions about data

What are the types of data that are especially relevant to L/L?

What specialized visualizations are especially relevant to L/L data?

How important are data uncertainty and data errors to the users?

Research questions about data

What are the types of data that are especially relevant to L/L?

What specialized visualizations are especially relevant to L/L data?

How important are data uncertainty and data errors to the users?

How can we visualize confidence/error information, if we have it?

Research questions about data

What are the types of data that are especially relevant to L/L?

What specialized visualizations are especially relevant to L/L data?

How important are data uncertainty and data errors to the users?

How can we visualize confidence/error information, if we have it?

What kinds of mismatches are there between the original data models and the visualization data models?

Task typologies

Task	Proposed by		
	Shneiderman	Keim	Yi
Overview	✓		
Zoom, Abstract/Elaborate	✓	✓	✓
Details-on-demand	✓	✓	
Filter, Select,	✓	✓	✓
Relate, Connect,	✓	✓	✓
Extract,	✓		
Explore,			✓
Reconfigure			✓
Encode,			✓
History	✓		

Sources: Shneiderman 1996, Keim et al. 2006, Yi et al. 2007, Unsworth 2000; Many Eyes

Task typologies

Task	Proposed by			
	Shneiderman	**Keim**	**Yi**	***Unsworth***
Overview	✓			
Zoom, Abstract/Elaborate	✓	✓	✓	
Details-on-demand	✓	✓		
Filter, Select, *Selection*	✓	✓	✓	✓
Relate, Connect, *Comparing*	✓	✓	✓	✓
Extract, *Sampling*	✓			✓
Explore, *Discovering*			✓	✓
Reconfigure			✓	
Encode, *Representing*			✓	✓
History	✓			
Annotating				✓
Referring, linking				✓

Sources: Shneiderman 1996, Keim et al. 2006, Yi et al. 2007, Unsworth 2000; Many Eyes

Research questions about tasks

What other L/L relevant tasks are there?

References: Bamman et al. 2007, Passarotti 2013

Research questions about tasks

What other L/L relevant tasks are there?

e.g.

- Find typical/unusual data
- Data modification

References: Bamman et al. 2007, Passarotti 2013

Research questions about tasks

What other L/L relevant tasks are there?

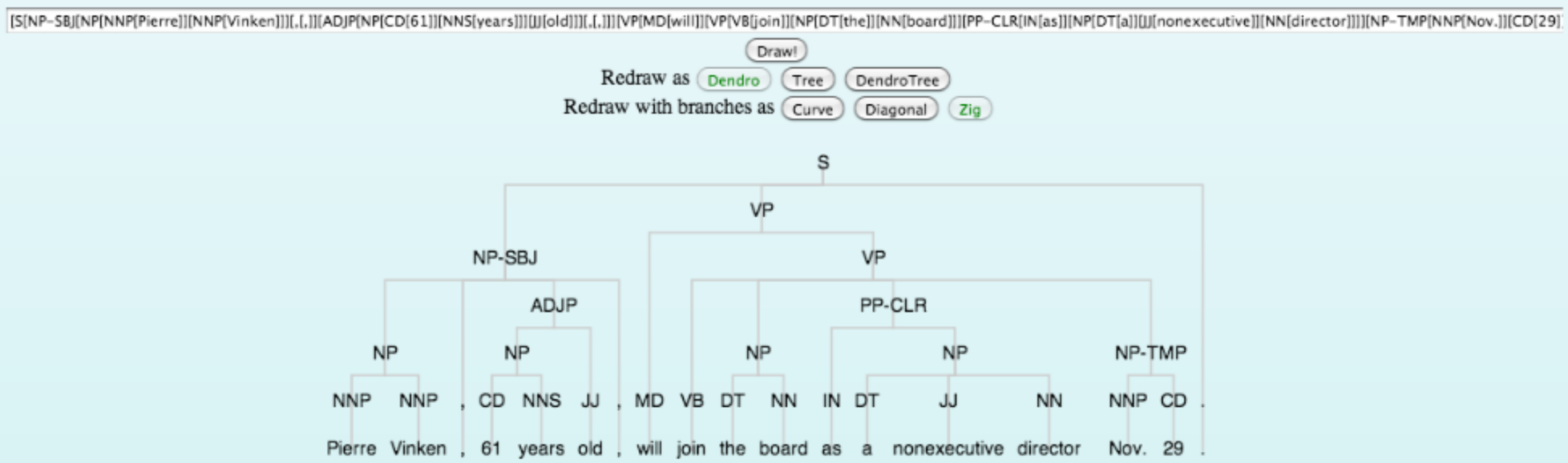
e.g.

- Find typical/unusual data
- Data modification

What tasks are non-L/L users interested in?

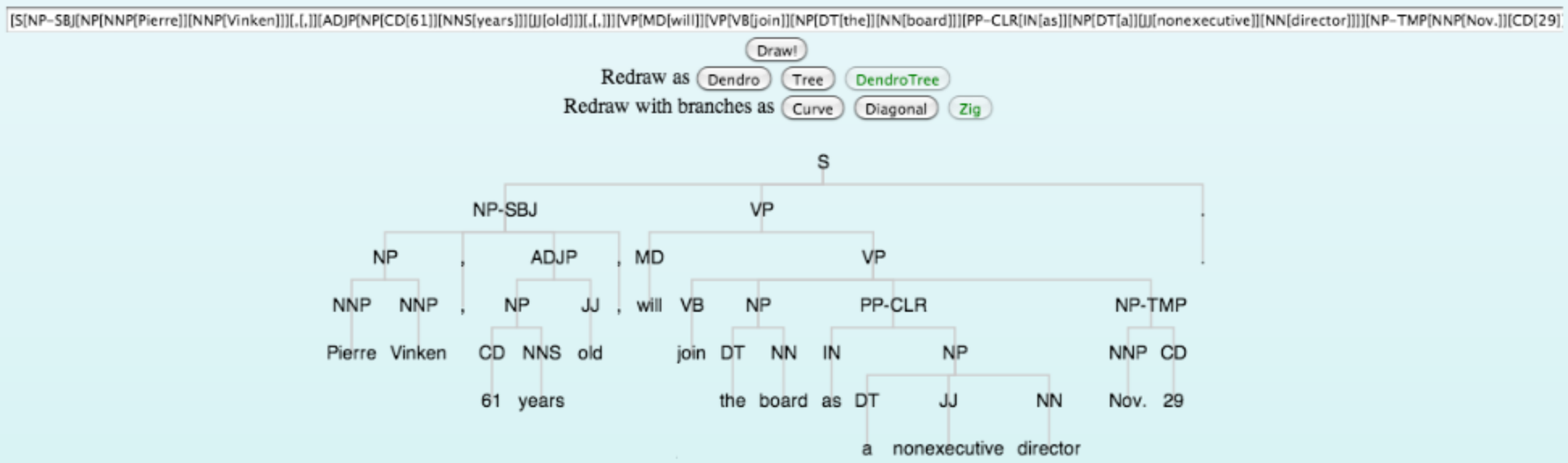
References: Bamman et al. 2007, Passarotti 2013

Which visualization for task, user?



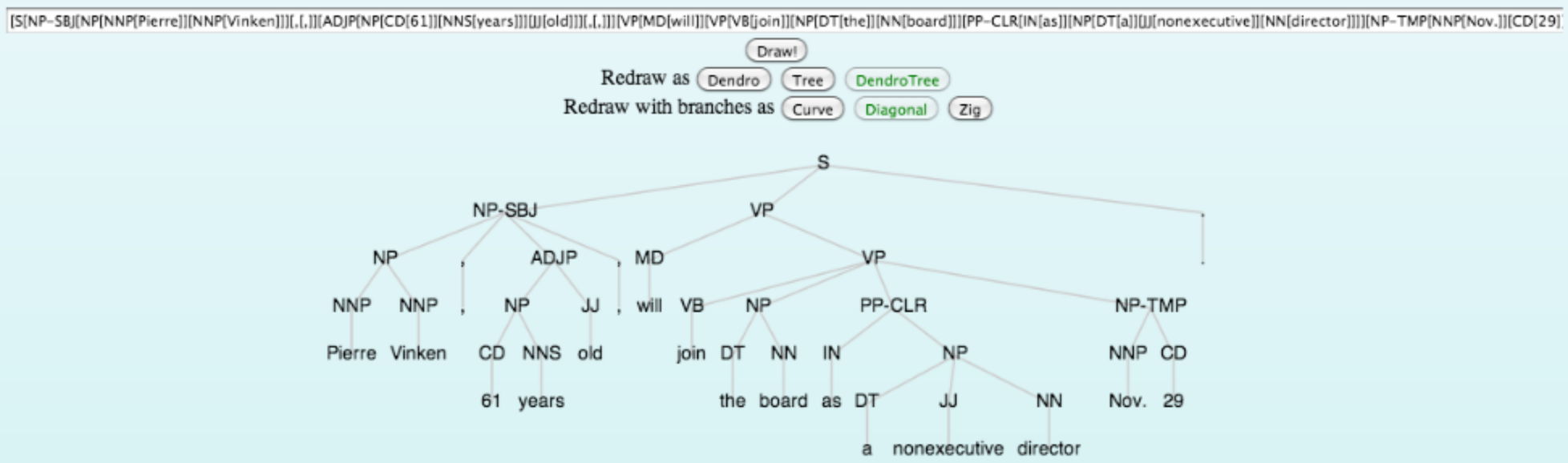
Uses: ProD

Which visualization for task, user?



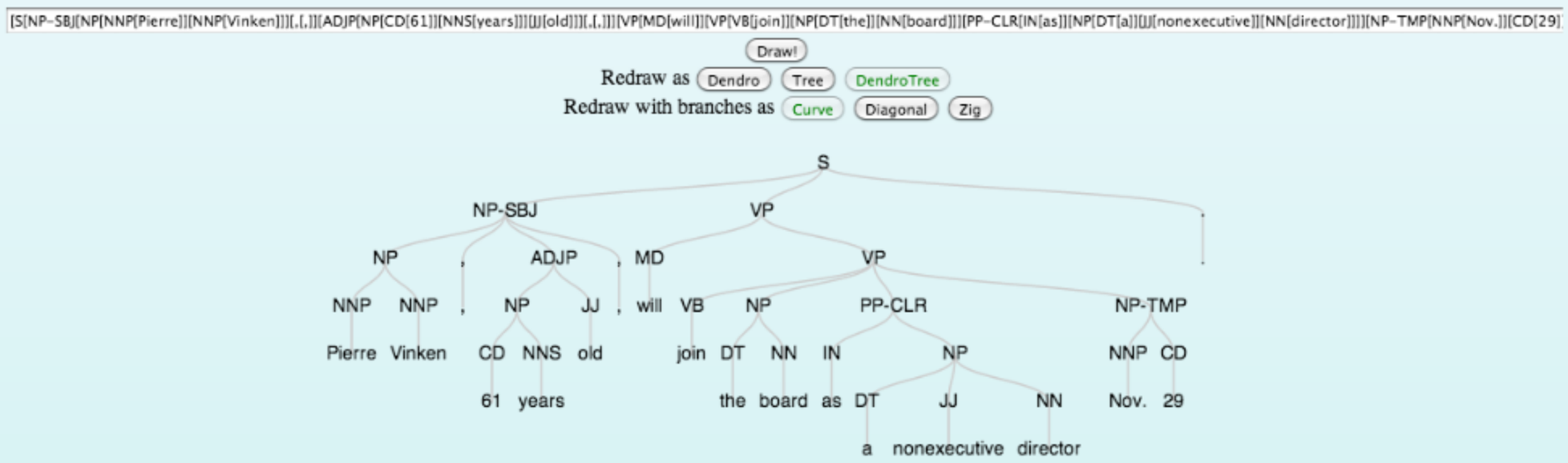
Uses: ProD

Which visualization for task, user?



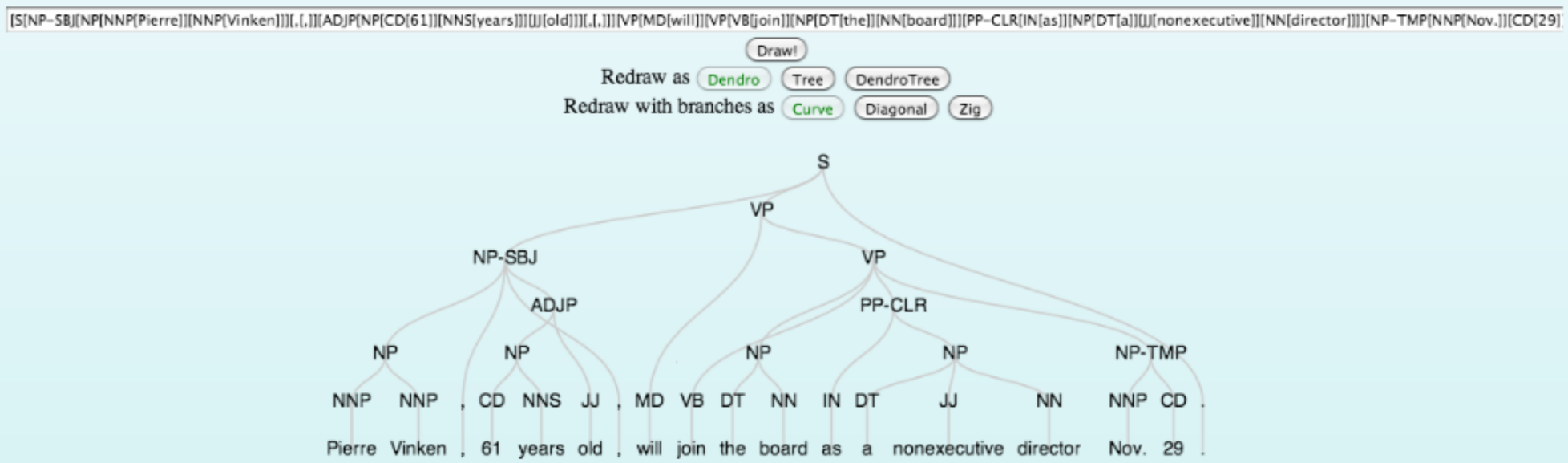
Uses: ProD

Which visualization for task, user?



Uses: ProD

Which visualization for task, user?



Uses: ProD

Which visualization for task, user?

Which visualization aspects are primarily and secondarily user preferences?

Which visualization for task, user?

Which visualization aspects are primarily and secondarily user preferences?

How are user preferences handled?

Uses: ProD

Resusability: generalizability

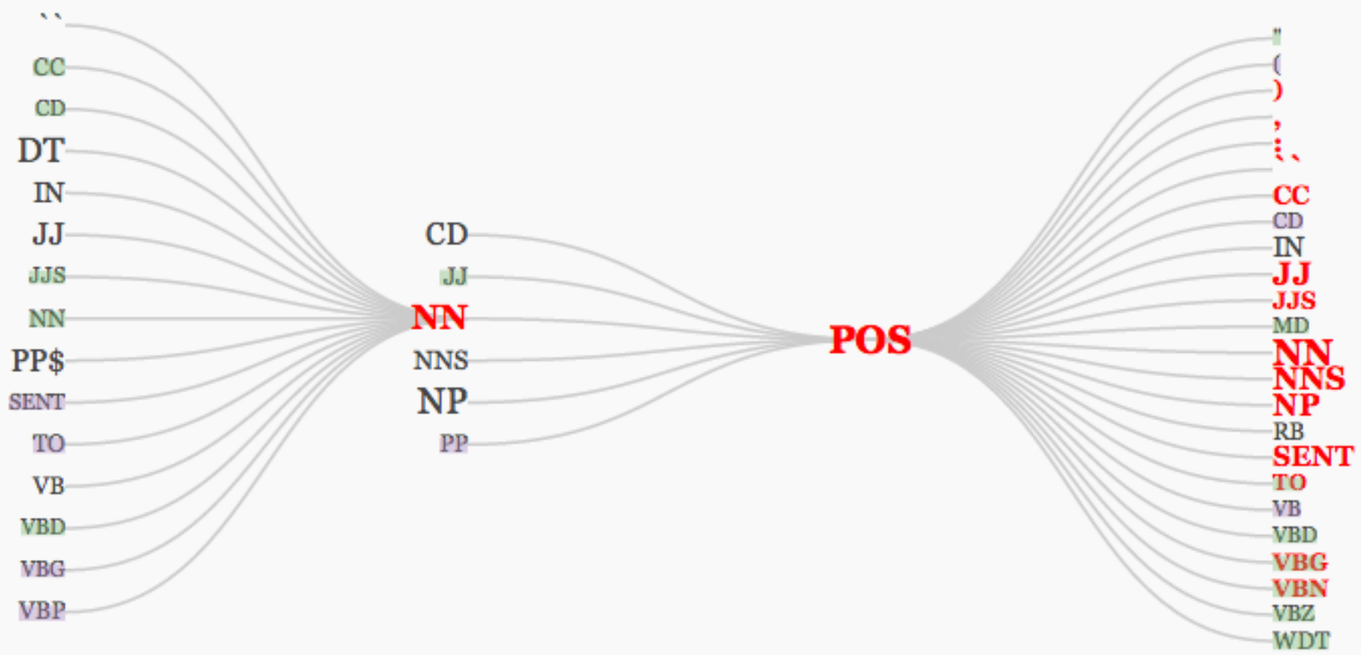
Input file: Choose File No file chosen POS to use: pos

Sort branches by: POS branching

Limit first left POS to:

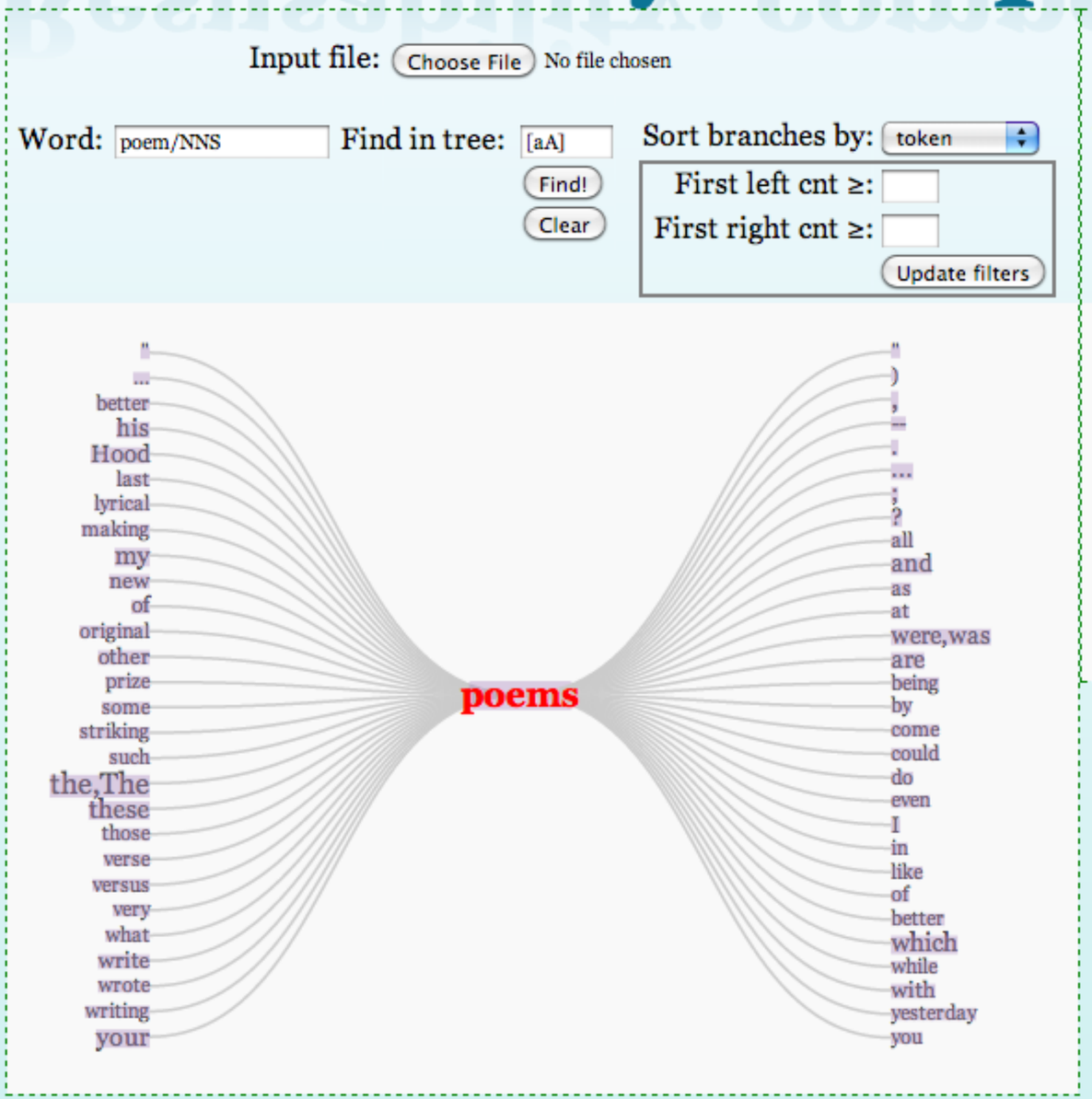
Limit first right POS to: Update filters

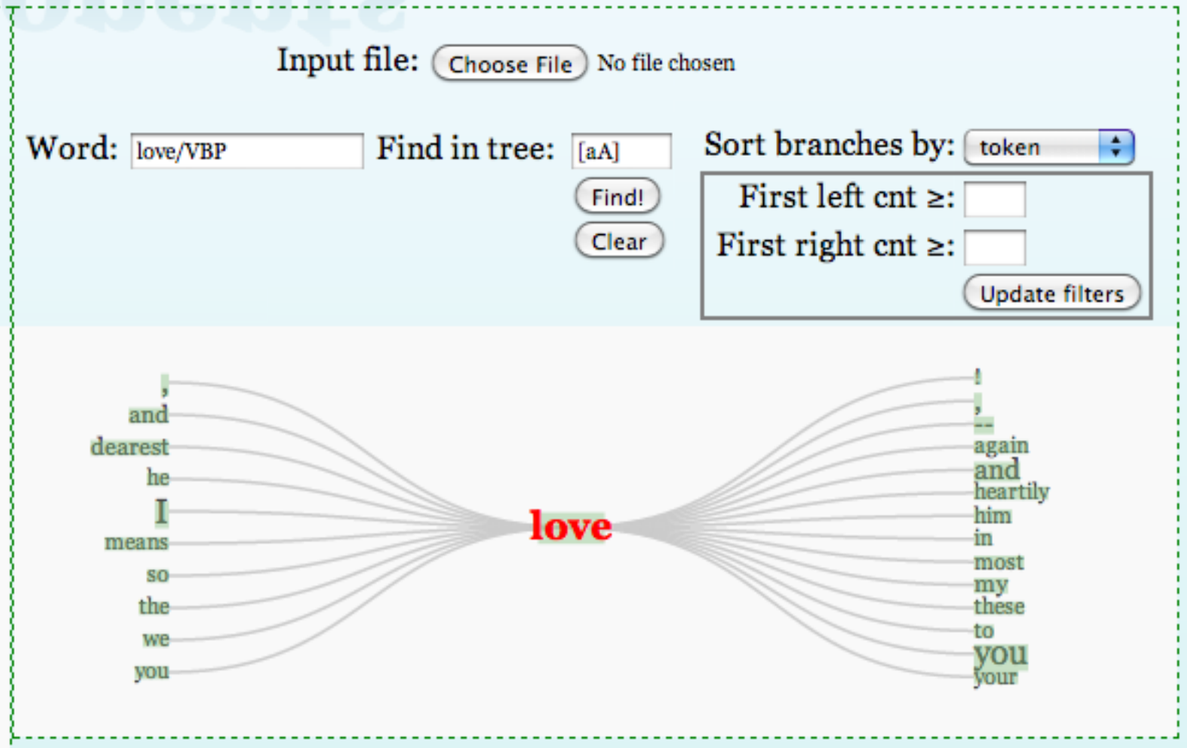
Find in tree: [aA] Clear



Uses: DoubleTreeJS

Resusability: components





Uses: DoubleTreeJS

Research questions about reusability

What are the (abstract) data properties that make a given visualization relevant for the data?

References: Grinstein's *WEAVE*, Stasko's *Jigsaw*

Research questions about reusability

What are the (abstract) data properties that make a given visualization relevant for the data?

What are the fundamental properties and actions of visualizations that form the basis for reusable components?

References: Grinstein's *WEAVE*, Stasko's *Jigsaw*

Visualizations put ideas into our heads!

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

PHILOSOPHISCHE
FAKULTÄT
Seminar für Sprachwissenschaft

Thank You

christopher.culy@uni-tuebingen.de

http://www.sfs.uni-tuebingen.de/~cculy/