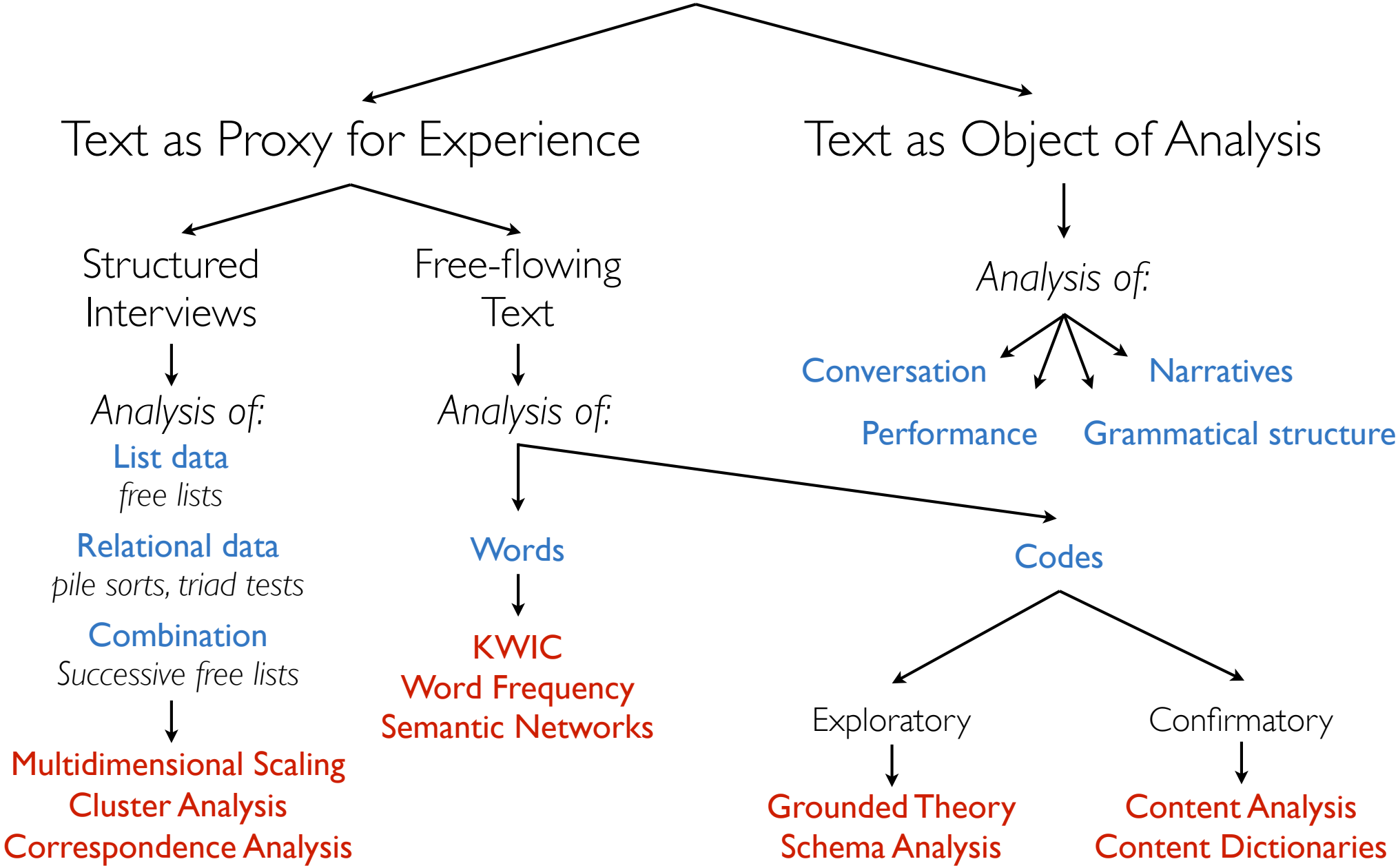
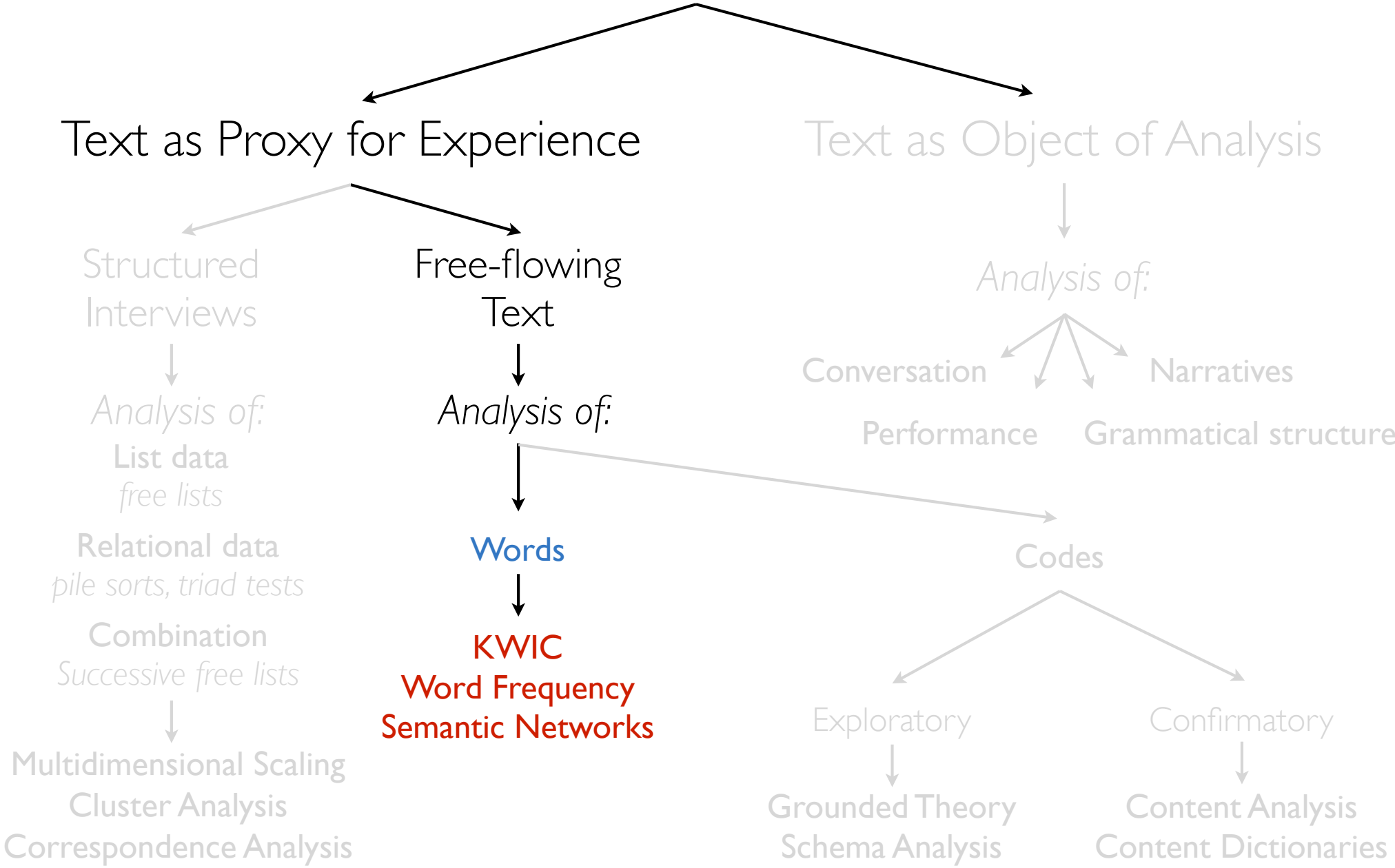
Traditions of Text Analysis



Adapted from Gery Ryan and H. Russell Bernard (2000)

Traditions of Text Analysis



Adapted from Gery Ryan and H. Russell Bernard (2000)

Key-Words-In-Context (KWIC)

			ID	Source
a violinist of the highest caliber	PLAY	a concerto. Jimmy has been taking	5D1	Williamson undergrad
leadership, that Scott will probably	PLAY	a larger role in this area, and	2H4	Shapiro small group
best buddy... how about coming over to	PLAY	a little ball...his friend	5D1	Williamson undergrad
while he is young he should learn to	PLAY	a musical instrument, the violin	5D1	Williamson undergrad
a great deal to us) could conceivably	PLAY	a significant part. It surprises me	2H5	Shapiro small group
an air of disdain. How can he possibly	PLAY	a solo in the concert tomorrow	5D1	Williamson undergrad
come to play a tune I have come to	PLAY	a song. Will you give me your	0A3	Zinacantan dreams
parents have decided that he should	PLAY	an instrument and have bought him	5D1	Williamson undergrad
well grown they used to go outside to	PLAY	and have fresh air. They went on in	8J1	Kikuyu
things, the children would ask her to	PLAY	and that if she sees that it is	1A1	Fam interaction
boy. When I was down south, I used to	PLAY	and then I would get mad and go in	5A2	Hartman Detroit gangs
we wanted something, and so we would	PLAY	and throw boxes all over the room,	5A2	Hartman Detroit gangs
you see I will be back I am going to	PLAY	at the house of our friend Manvel	0A3	Zinacantan drea,ms
and scold him and tell him he cannot	PLAY	ball at all today since he did not	5D1	Williamson undergrad
is a nice spring day and he wanted to	PLAY	ball instead. He is now trying to	5D1	Williamson undergrad
he is in the attic, watching some boys	PLAY	baseball on the opposite hill. He	5D1	Williamson undergrad

[Stone et al. 1966:159]

Word Frequency Example

- Goal: Are there differences between how mothers and fathers describe their children?

- Data: Weisner & colleagues asked 82 mothers and 82 fathers "What is your teenager like now? Does she or he have any special qualities or abilities?"

- Examples of parents' descriptions of their children .
  - ID009. F1030. Boy. Fthr. Loving. Obedient. Maintains own identity. Likes being home. Independent. Anxious to go to California to school.

  - ID124. F1130. Girl. Mthr. Great kid. Willing to communicate with parents. Listens. Motivated in school. Helpful around the house. Healthy. Active. Lots of friends. She tends to play it safe.

[Ryan & Weisner 1996]

(Note: Each quality is treated as a sentence)

Word List Example

What is your teenager like now? Does she or he have any special qualities or abilities?

Mothers

Rank	Frequency	Word
1	22	good
2	12	friends
3	11	loving
4	11	out
5	11	people
6	10	doesn't
7	10	hard
8	10	school
9	9	responsible
10	9	sense
11	8	caring
12	8	intelligent
13	8	lacks
14	8	sensitive
15	7	bright
16	7	honest
17	7	others
18	7	self
19	7	time
20	7	well
21	7	work
...	...	...
666	1	zest

Fathers

Rank	Frequency	Word
1	22	good
2	16	school
3	11	hard
4	9	intelligent
5	8	bright
6	8	independent
7	8	out
8	8	well
9	7	doesn't
10	7	lack
11	7	loving
12	7	people
13	7	sensitive
14	7	sports
15	7	student
...	...	...
548	1	zero

Ryan and Weissner (1996)

Word Frequency Example
(Standardized Differences)

WORD	Both	Mother	Father		Expected Father	Difference (Exp. F - M)
school	26	10	16	* 1.27 =	20.3	10.3
good	45	22	23		29.2	7.2
lack	9	2	7		8.9	6.9
student	9	2	7		8.9	6.9
enjoys	6	1	5		6.4	5.4
independent	13	5	8		10.2	5.2
great	11	6	5		6.4	0.4
mature	11	6	5		6.4	0.4
humor	9	5	4		5.1	0.1
times	9	5	4		5.1	0.1
attitude	7	4	3		3.8	-0.2
caring	14	8	6		7.6	-0.4
honest	9	7	2		2.5	-4.5
time	9	7	2		2.5	-4.5
creative	6	6	0		0.0	-6.0
friends	16	12	4		5.1	-6.9

[Ryan & Weisner 1996]

Cultural Model of *Color* in Puerto Rico

- What are the salient concepts and categories that constitute the cultural model of *color*?
- How is meaning organized according to this cultural model?
- How shared is the cultural model of *color* across age, sex, class, and color?

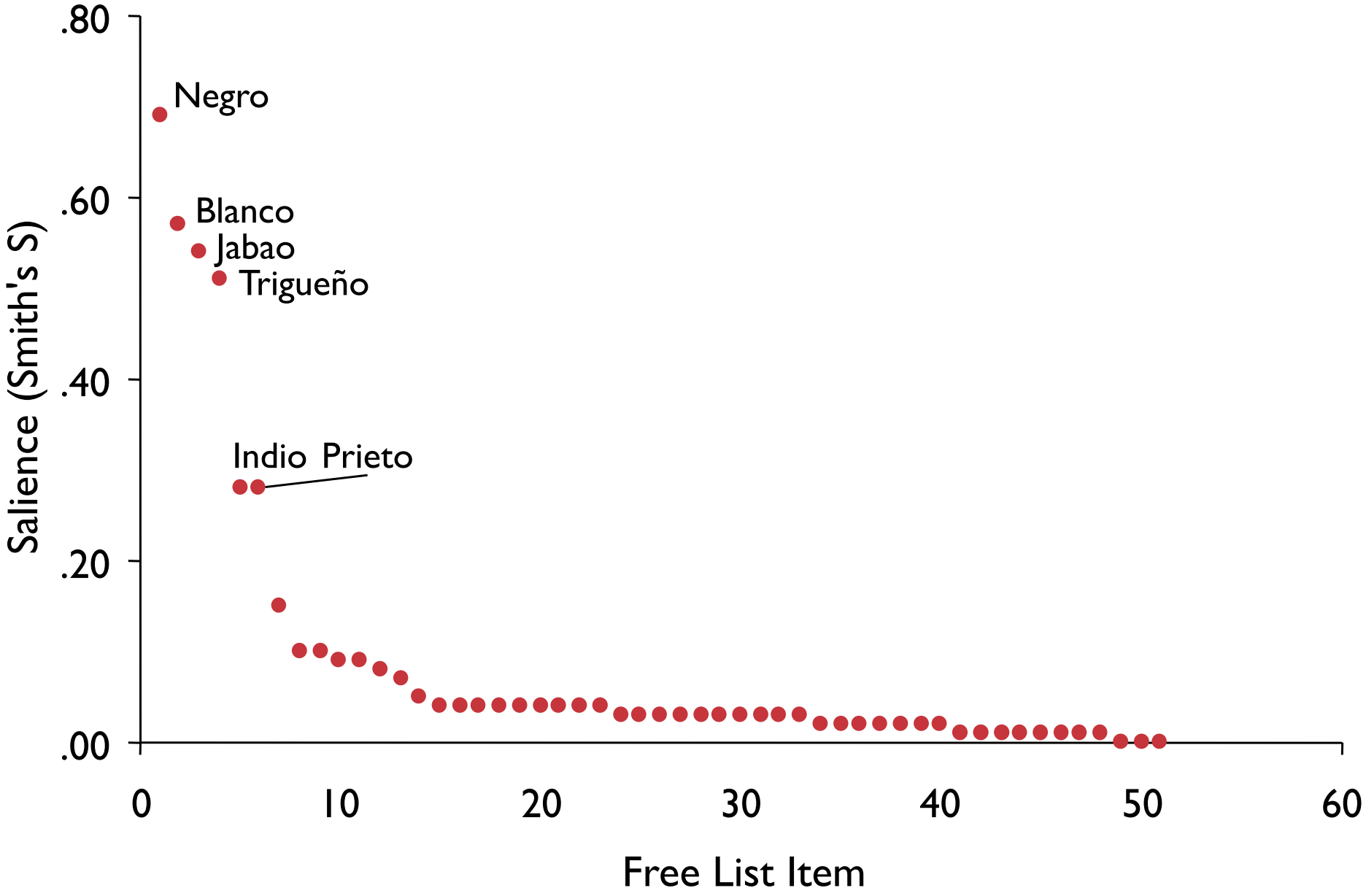
“Aquí no hay divisiones de raza, pero sí de color”

“Maria,” 32-year old shopkeeper

When I refer to *raza*, I'm referring to *la raza puertorriqueña*… I'm not referring to colors. I'm talking there about people in general. Here there aren't divisions of *raza*, but there are divisions of *color*.

CG: Alright, explain that to me.

OK. The Puerto Rican *race* is white and black, *jabao* [roughly, high yellow], blonde, dark, *trigueño* [literally, "wheat-colored"], as people say. It's a single *raza*. We have just a single *raza*. And that's how we see ourselves. We're the Puerto Rican *raza*. But within the Puerto Rican *raza*, there are the blacks, the high yellows, the *indios* [literally, "Indians"], and the *trigueños*, as many people tell you. What many people call *trigueños* are *indios*.

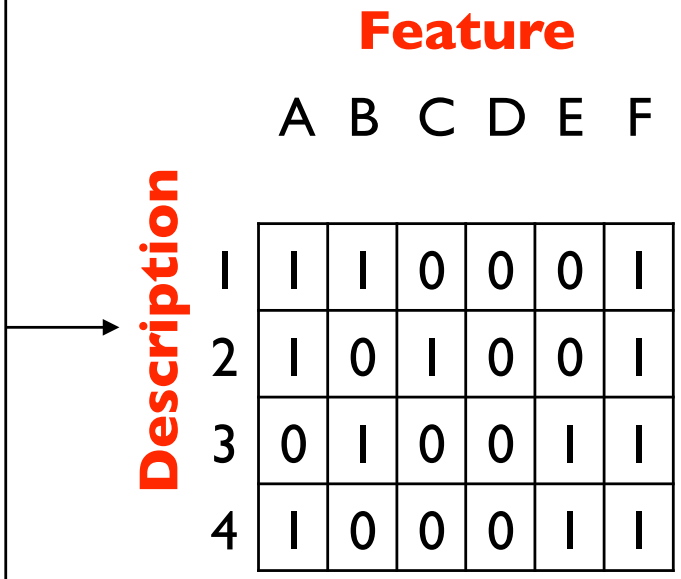


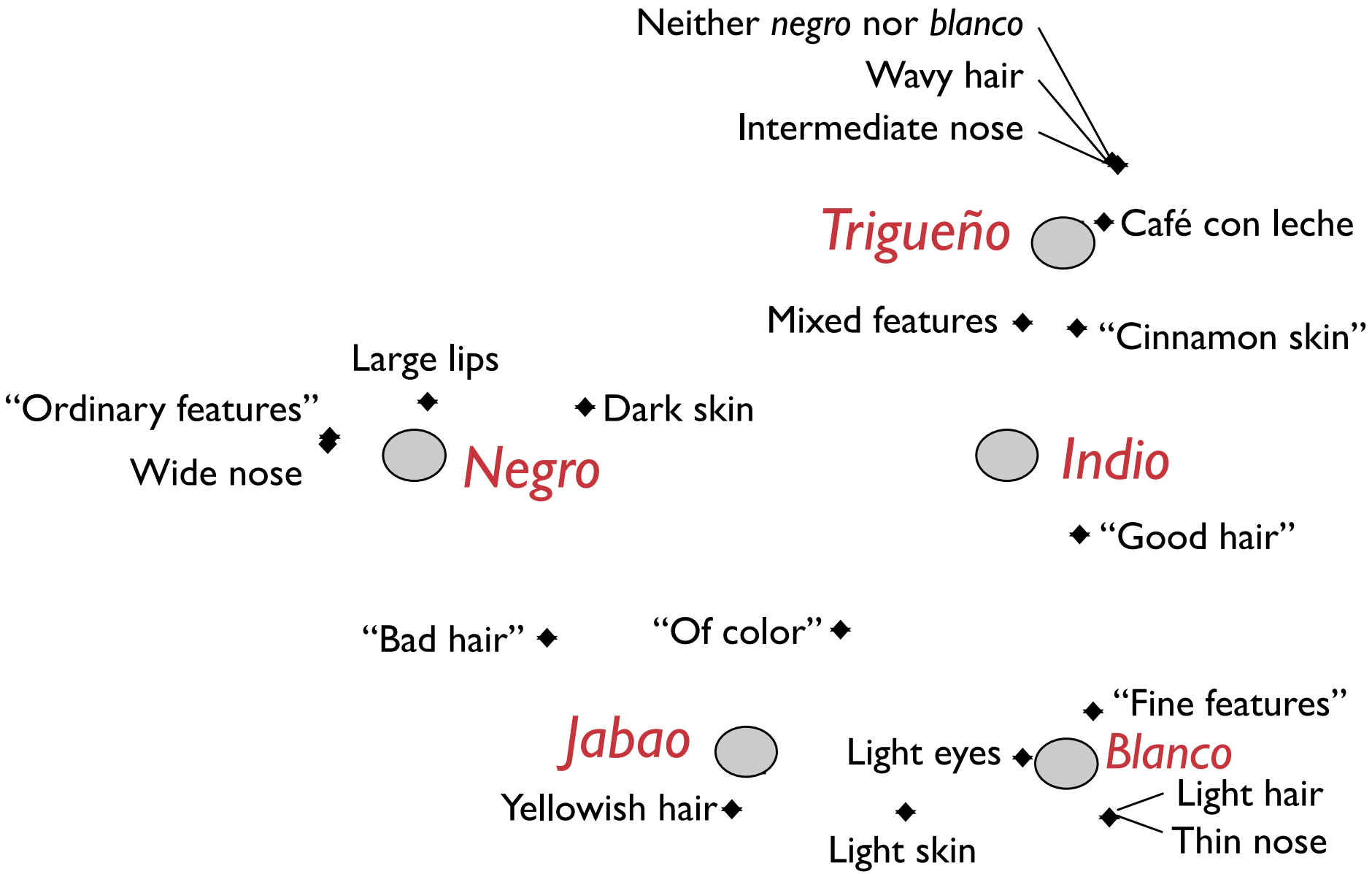
Gravlee, C. C. 2005. *Social Forces* 83:949-970.

Successive Free Listing

- 13 informants described the characteristics associated with each term in their free list of *colores*
- Result was 98 open-ended descriptions of 27 emic *color* terms
- Successive free listing produces 2-mode relational data
- Can visualize relations with correspondence analysis

So then we have the *jabao*. The *jabao* is supposedly a combination that comes from a *blanco* and a *negro*. It could be someone really black who mixed with a white person, married a white person, and then. . .out comes this person with really light skin. He's white, in effect, but son of a black man and a white woman. His hair might be bad [kinky], what we call "crispy," but of a different color. A half-yellow color that some blondes have, but it's bad hair. The hair's bad. His eyes might be light-colored, even green or blue. That's what we call a *jabao* here. The skin is light, the guy is light. The eyes can be black or green or whatever. But his skin is light, and his hair always tends to be "crispy."




Neither *negro* nor *blanco*
Wavy hair
Intermediate nose
Trigueño
Café con leche
Mixed features
"Cinnamon skin"
Large lips
"Ordinary features"
Dark skin
Wide nose
Negro
Indio
"Good hair"
"Bad hair"
"Of color"
"Fine features"
Jabao
Light eyes
Blanco
Light hair
Yellowish hair
Thin nose
Light skin

“Someone who is white but has bad hair, they call that person *jabao*, because he’s really white like you, but his hair is curly. That’s a *jabao*.

CG: If I had bad hair—

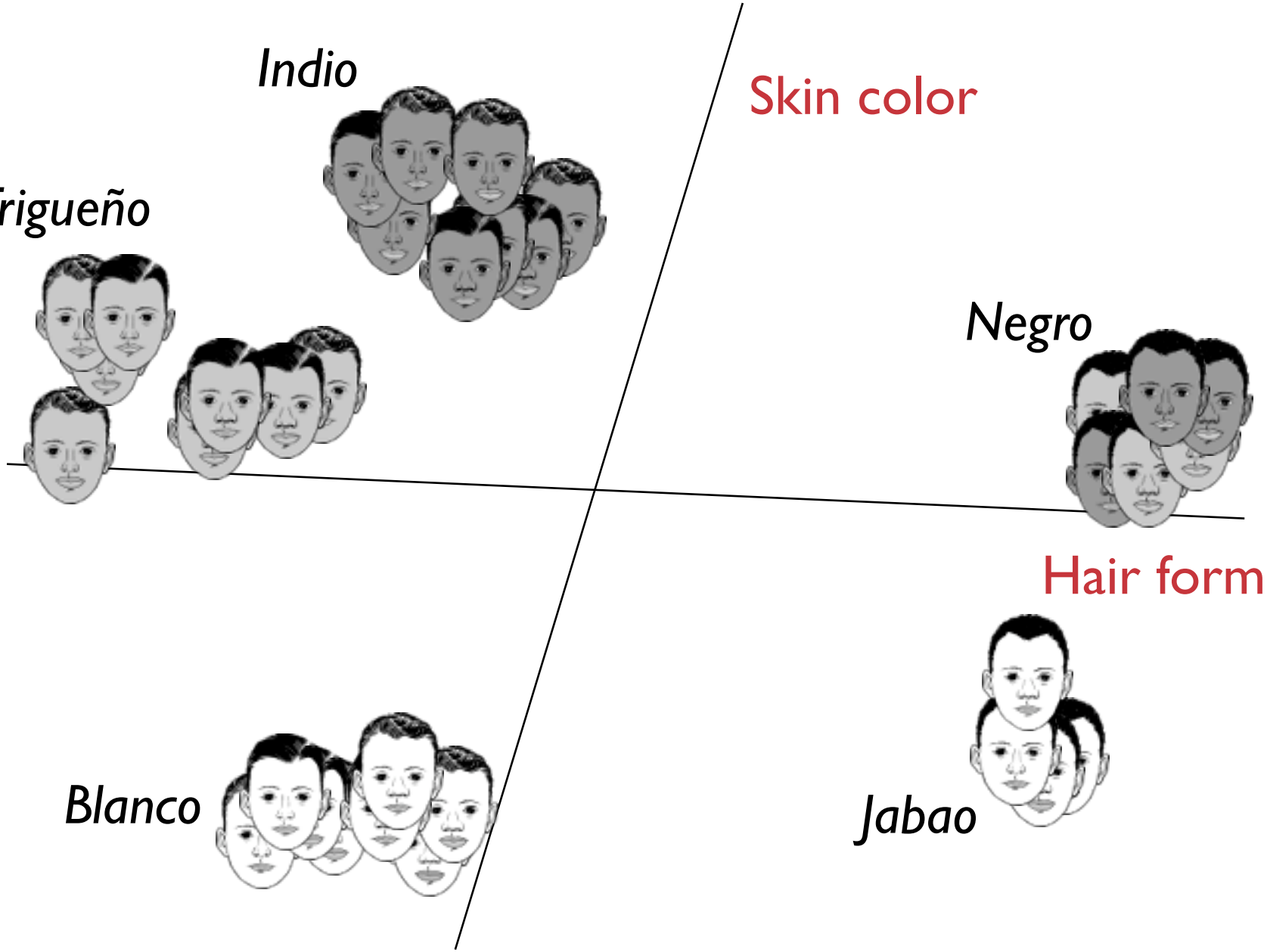
You’d be *jabao*. That is, if you got the bad hair from your father or mother, one of them has part of the black race.”

“Don Fernando”
75-year old former mechanic

"I think it's the hair. For me, it's the hair. With the trigueño, she's trigueña [here: dark-skinned], but her hair is like yours. With the negro, she's of color [dark-skinned], but her hair, well she has to go to the beauty parlor to get her hair—to put her hair like she wears it. It's the type of hair that's all curled up. The hair that has to be straightened."

"Doña Diana"

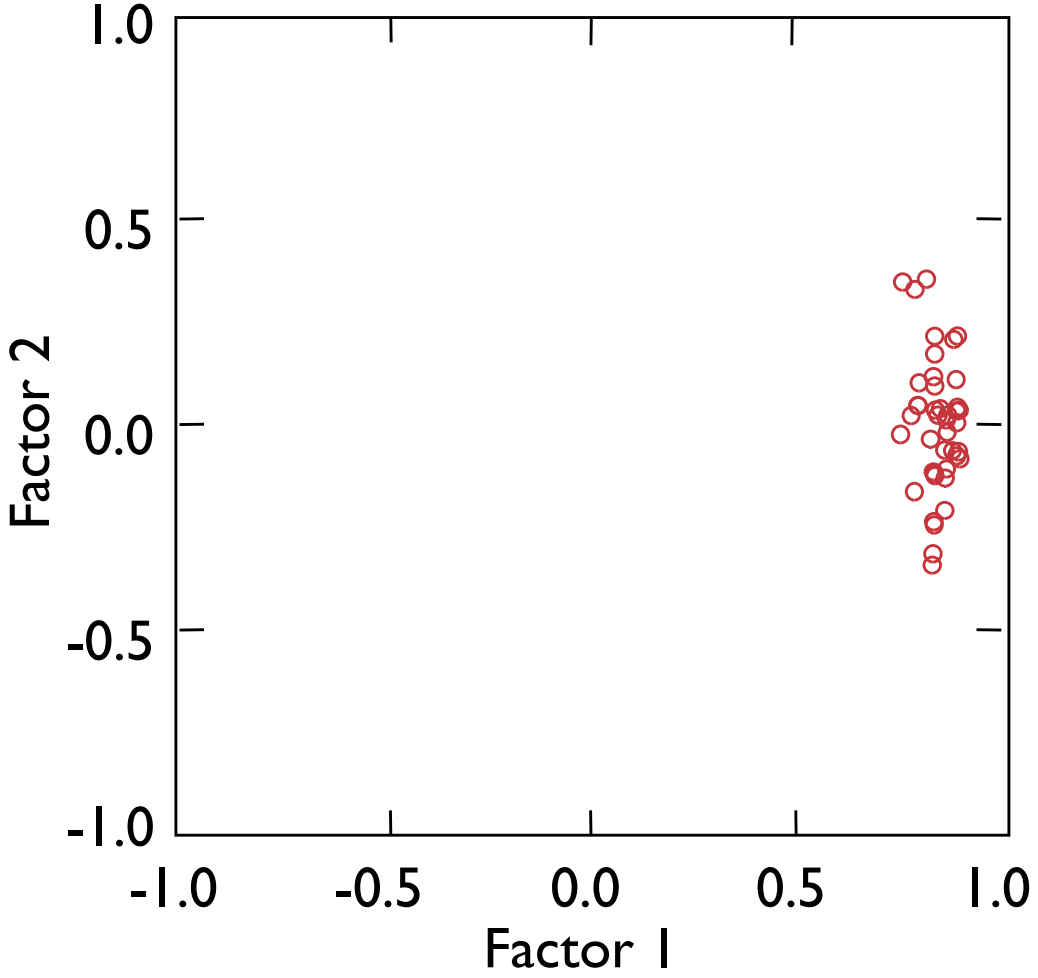
53-year old nurse



Stress = .186

Gravlee, C. C. 2005. *Social Forces* 83:949-970.

Single, Shared Cultural Model



71.7%

variance explained

27.9

factor 1:factor 2

.84 ± .04

average factor loading

Gravlee, C. C. 2005. *Social Forces* 83:949-970.

(A)
Original Text

Respondent 1: I'm concerned about the agricultural acreage. I can't believe there is still 400000 acres of ag in maricopa county. Also, S&V water will not be used by the ag on reservation lands. The ag acreage will only go away within SRP and on non-reservation lands. There's just not that much left in MC. The land that will be developed between now and 2030 is desert land dependent upon CAGRD (CAP and reclaimed water). Does this factor in pumping from indians? ...

Respondent 2: Growth projections seem too conservative. We and most cities have found DES and Census data to be wrong based on field experience. The cost of water is a bigger demand factor than previously thought.

Respondent 3: I question the per capita demand projections. We are seeing declines due to increased use of low flow fixtures and xeriscape. The population projections using the density of development projections from MAG (we call it the "red dot" models) will help with the Ag retirement and phasing of development densities. You might also want to contact ADWR about in-state shortage sharing assumptions to get a better understanding about CAP reductions within Maricopa County in the various drought scenarios....

Respondent 4: It appears to be very appropriate. Certainly easy to understand and comprehend. Though not a technical expert, it appears as though using historical trends and anticipating population growth and changing of land uses is a good strategy. I also like that the option to change the DES forecasts for population growht is an option. Our community is experiencing growth that is proving to cause the current DES population figures to be highly underestimated.

(B)
Word-by-Respondent Profile Matrix

Word	Resp 1	Resp 2	Resp 3	Resp 4
be	2	1	0	2
projections	0	1	3	0
will	3	0	1	0
ag	3	0	1	0
population	0	0	1	3
growth	0	1	0	2
also	1	0	1	1
des	0	1	0	2
land	2	0	0	1
water	2	1	0	0

(C)
Dichotomized Profile Matrix

Word	Resp 1	Resp 2	Resp 3	Resp 4
be	1	1	0	1
projections	0	1	1	0
will	1	0	1	0
ag	1	0	1	0
population	0	0	1	1
growth	0	1	0	1
also	1	0	1	1
des	0	1	0	1
land	1	0	0	1
water	1	1	0	0

(D)
Respondent-by-Respondent Proximity Matrix

	Resp 1	Resp 2	Resp 3	Resp 4
Resp 1	1.00	0.20	0.20	0.30
Resp 2	0.20	1.00	0.10	0.30
Resp 3	0.20	0.10	1.00	0.20
Resp 4	0.30	0.30	0.00	0.00

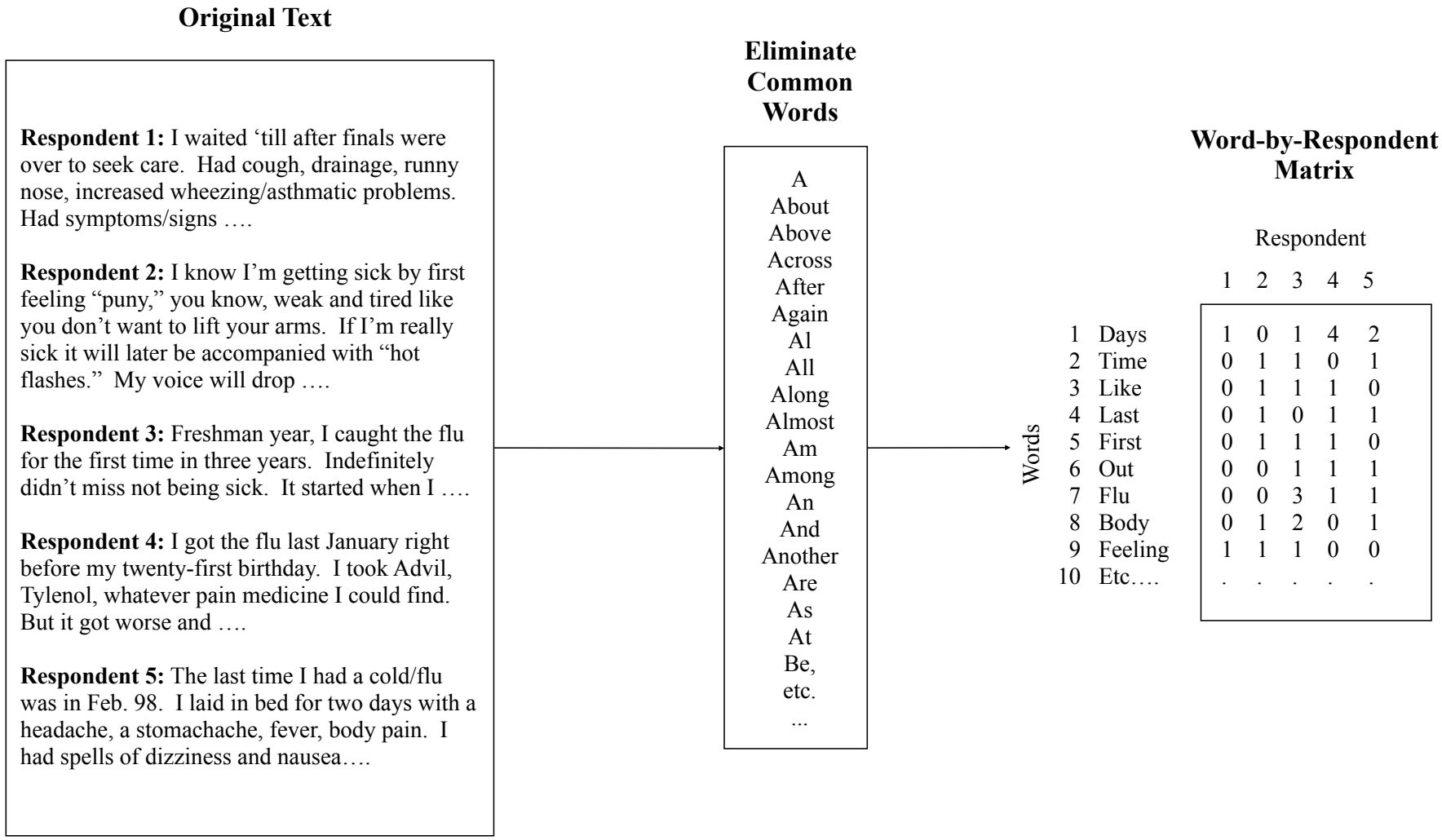
(E)
Word-by-Word Proximity Matrix

	be	projections	will	ag	population	...
be	1.00	0.25	0.25	0.25	0.25	...
projections	0.25	1.00	0.25	0.25	0.25	...
will	0.25	0.25	1.00	0.50	0.25	...
ag	0.25	0.25	0.50	1.00	0.25	...
population	0.25	0.25	0.25	0.25	1.00	...
...	...	...	...	...	...	...

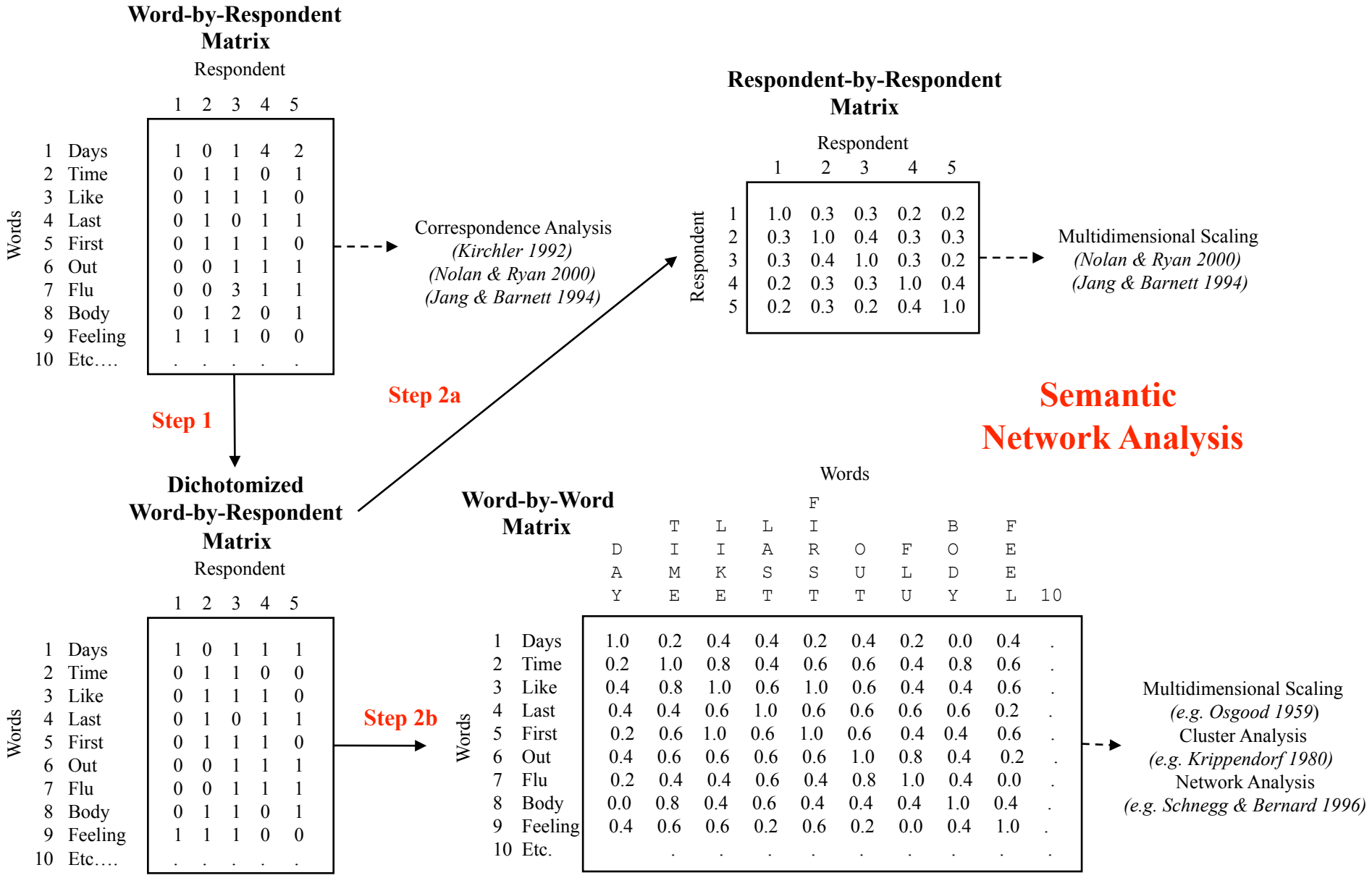
Wutich, A. & Gravlee C. C. (2010) In I. Vaccaro, E. A. Smith, S. Aswani (Eds.), Environmental Social Sciences: Methods and Research Design (in press). New York: Cambridge University Press

Fundamentals of Word-Based Text Analysis

Converting Texts to Matrices



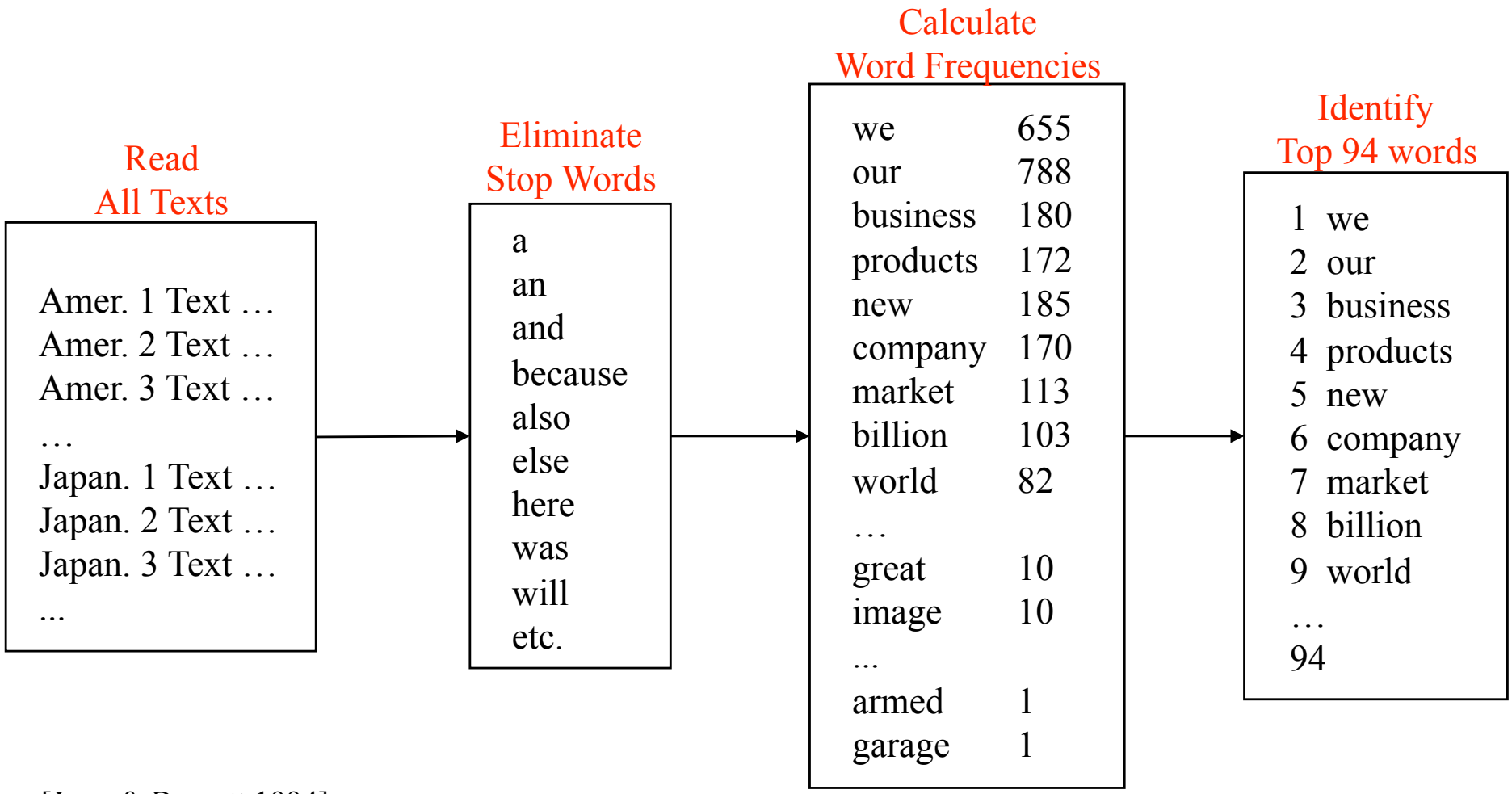
Semantic Network Analysis
(Analysis of Word-by-Respondent Text Matrices)



Semantic Network Example
(Text Management Steps)

Goal: Compare American and Japanese business practices

Data: CEOs' yearly letters to stockholders from 35 firms (1992)



[Jang & Barnett 1994]

Semantic Network Example

(Text Management Steps)

Top 94 words

1 we
2 our
3 business
4 products
5 new
6 company
7 market
8 billion
9 world
…
94

+

All Texts

Amer. 1 Text …
Amer. 2 Text …
Amer. 3 Text …
…

Japan. 1 Text …
Japan. 2 Text …
Japan. 3 Text …
…

Top 94 words

35 Companies	We	Our	Business	Products
A1	27	8	25	10
A2	9	1	1	8
A3	53	9	16	48
…				
J4	5	1	3	10
J5	12	2	1	6
J6	12	1	0	18
…	.	.	.	.

35 Companies

35 Companies	A1	A2	A3	J1	J2
A1	1.0	.89	.78	.43	.27
A2	.89	1.0	.62	.38	.56
A3	.78	.62	1.0	.23	.42
J1	.43.	.38	.23	1.0	.87
J2	.27	.56	.42	.87	1.0
…	.	.	.	.	.

Multidimensional Scaling

[Jang & Barnett 1994]

Semantic Network Example

(Multidimensional Scaling of Company-by-Company Matrix)

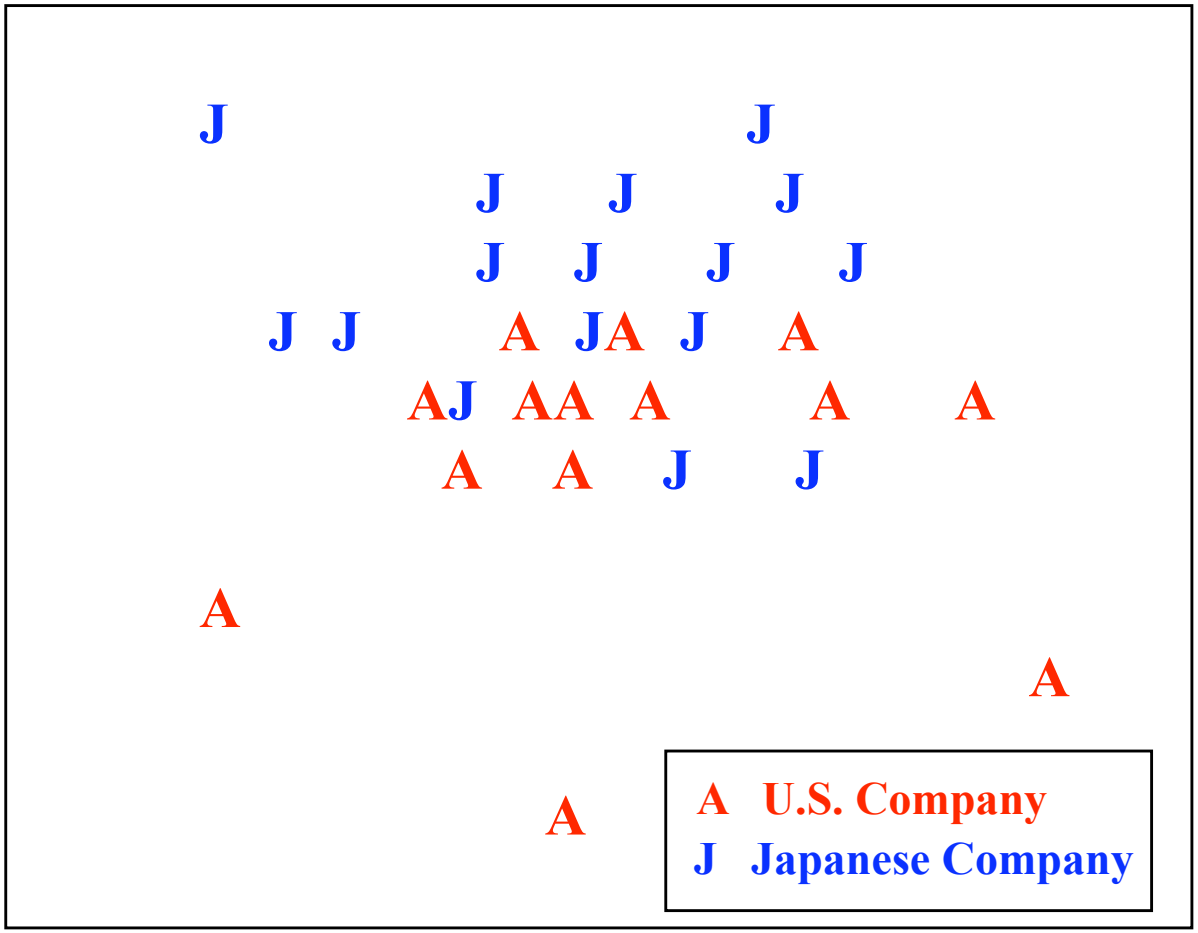
Companies

Top 95 Words	1	2	3	4	…
Income	8	8	3	3	
Major	9	13	23	12	
Success	31	32	5	28	
Improve	17	12	8	2	
Etc.	.	.	.	.	

Respondents (Companies)

Respondents (Companies)	A	B	C	D
A	-	.9	.3	.4
B	.9	-	.4	.2
C	.3	.4	-	.8
D	.4	.2	.8	-

MDS

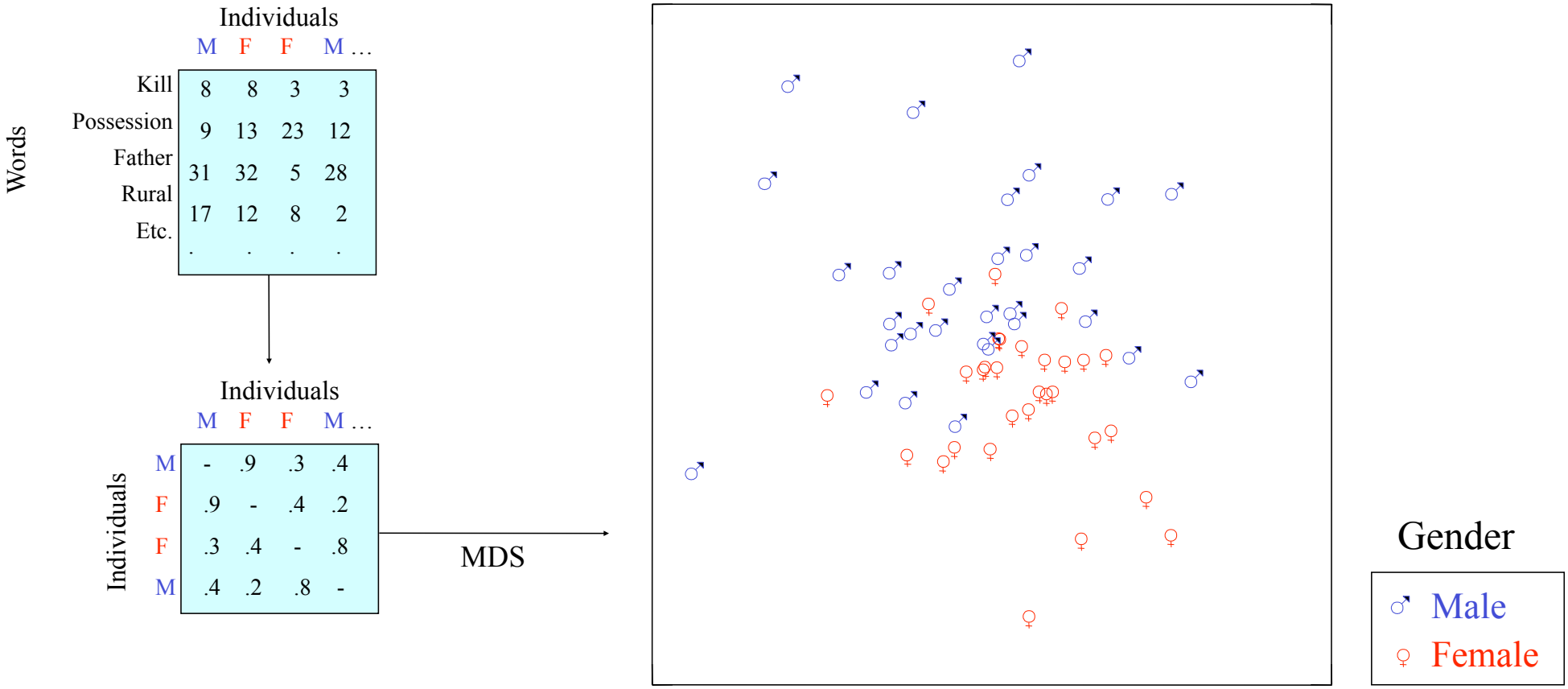


[Jang & Barnett 1994]

Undergraduates' Descriptions of 'Slasher' Films

- Nolan and Ryan, 2000
- 60 American college students (30 M/30 F)
- Most frightening film they could recall
- Describe emotional reactions to film
- Most popular: Halloween (9), Texas Chainsaw Massacre (4), Scream (4), Psycho (4), Friday the 13th Pt. I (4)

Multidimensional scaling of informants based on words used in descriptions of horror films (Nolan & Ryan 2000).



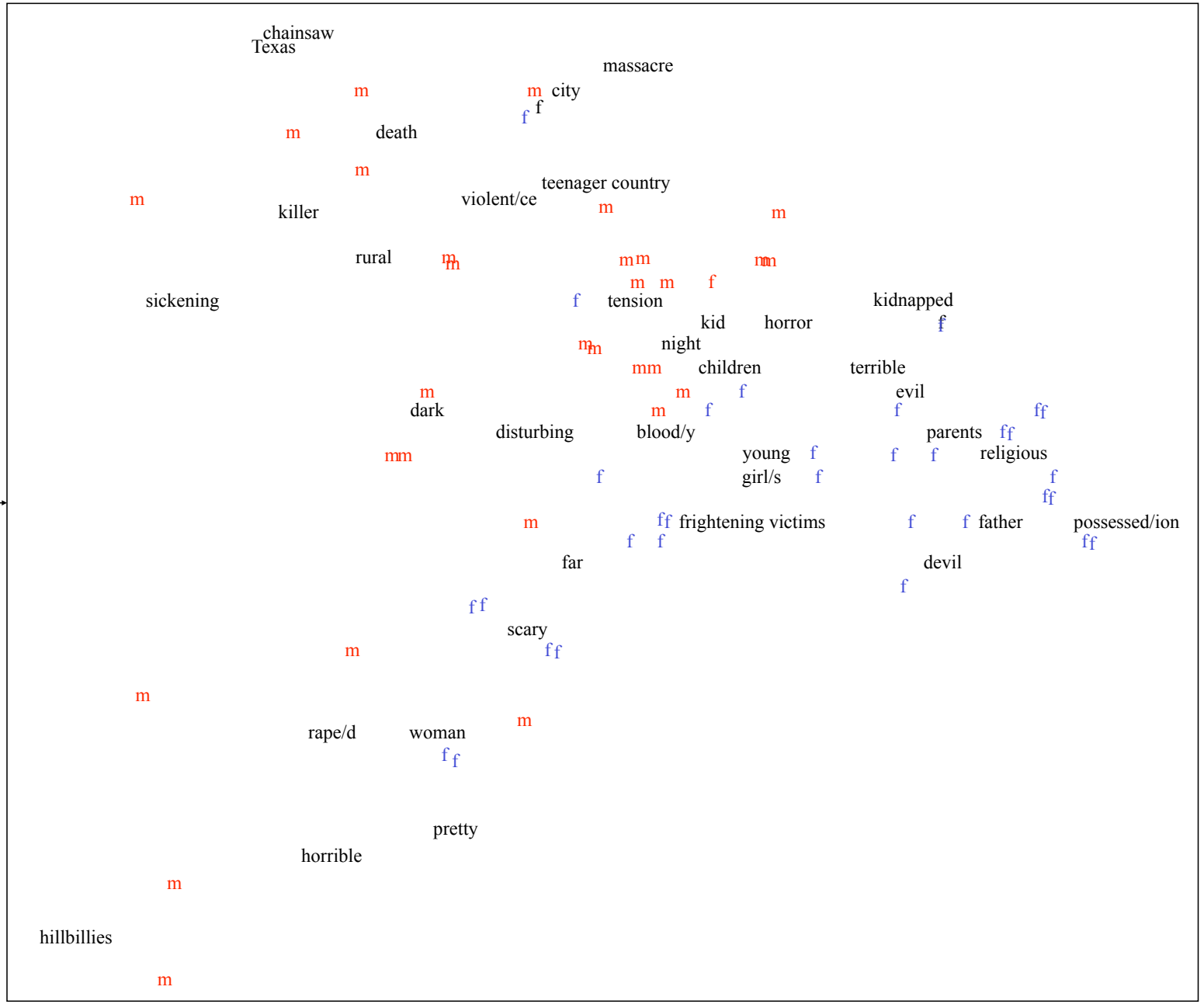
From Gery Ryan, RAND, 2006

Example of Word-Co-Occurrence (Correspondence Analysis of Word-by-Informant Matrix)

Individuals

Words	M	F	F	M …
killer	8	8	3	3
devil	9	13	23	12
parents	31	32	5	28
woman	17	12	8	2
Etc.	.	.	.	.

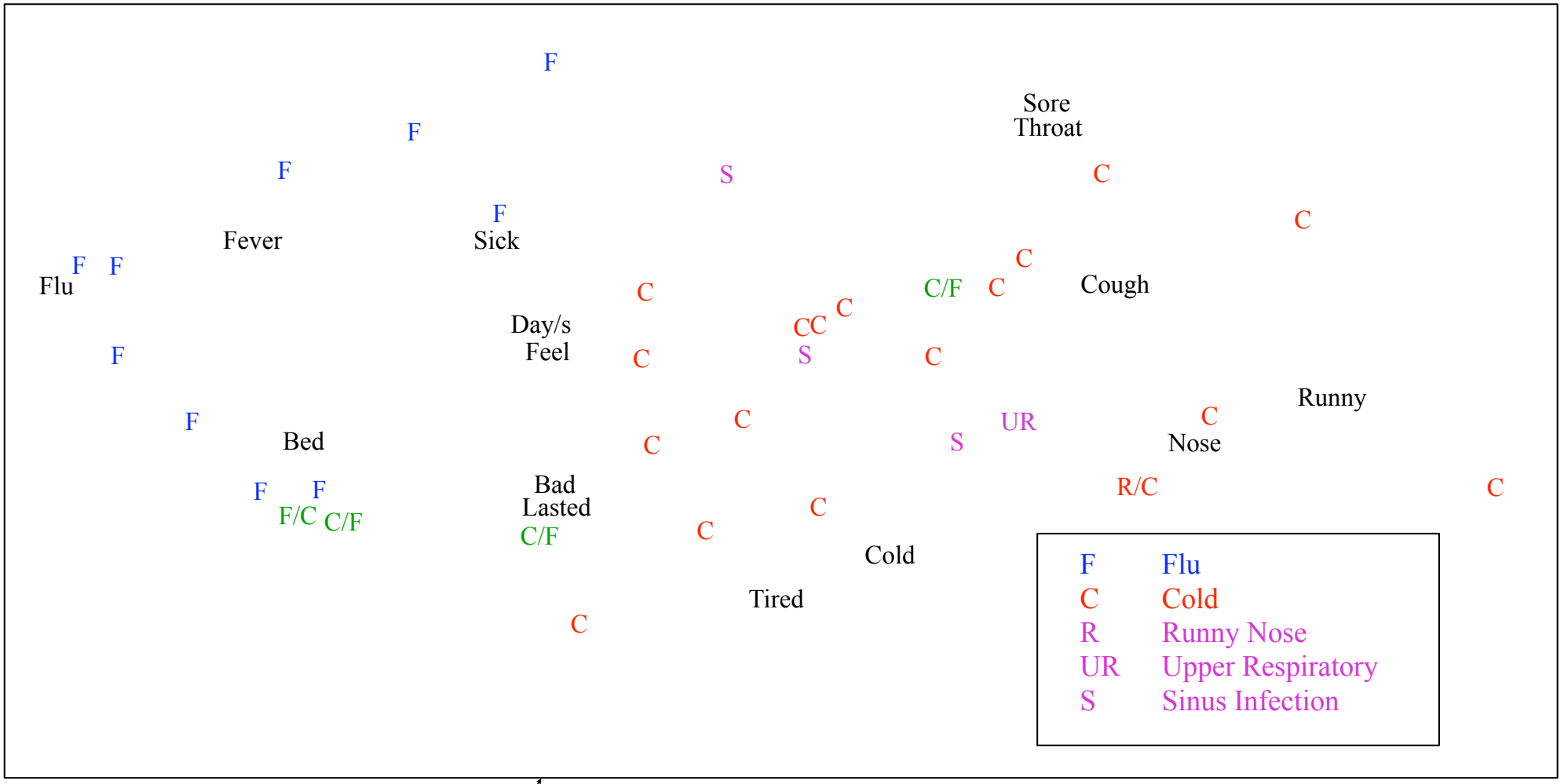
Correspondence Analysis

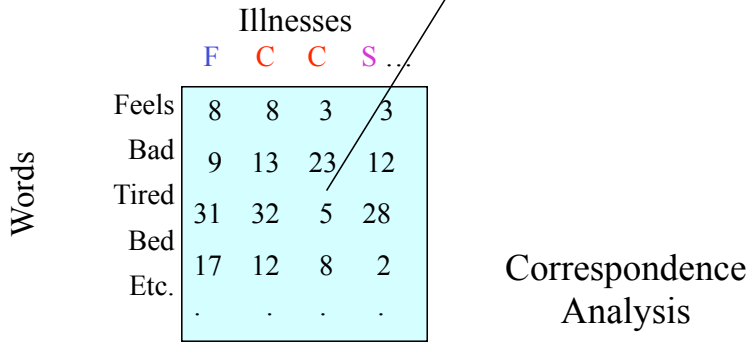


[Nolan & Ryan 2000]

From Gery Ryan, RAND, 2006

Semantic Network of Key Words and Self Diagnoses





Correspondence Analysis