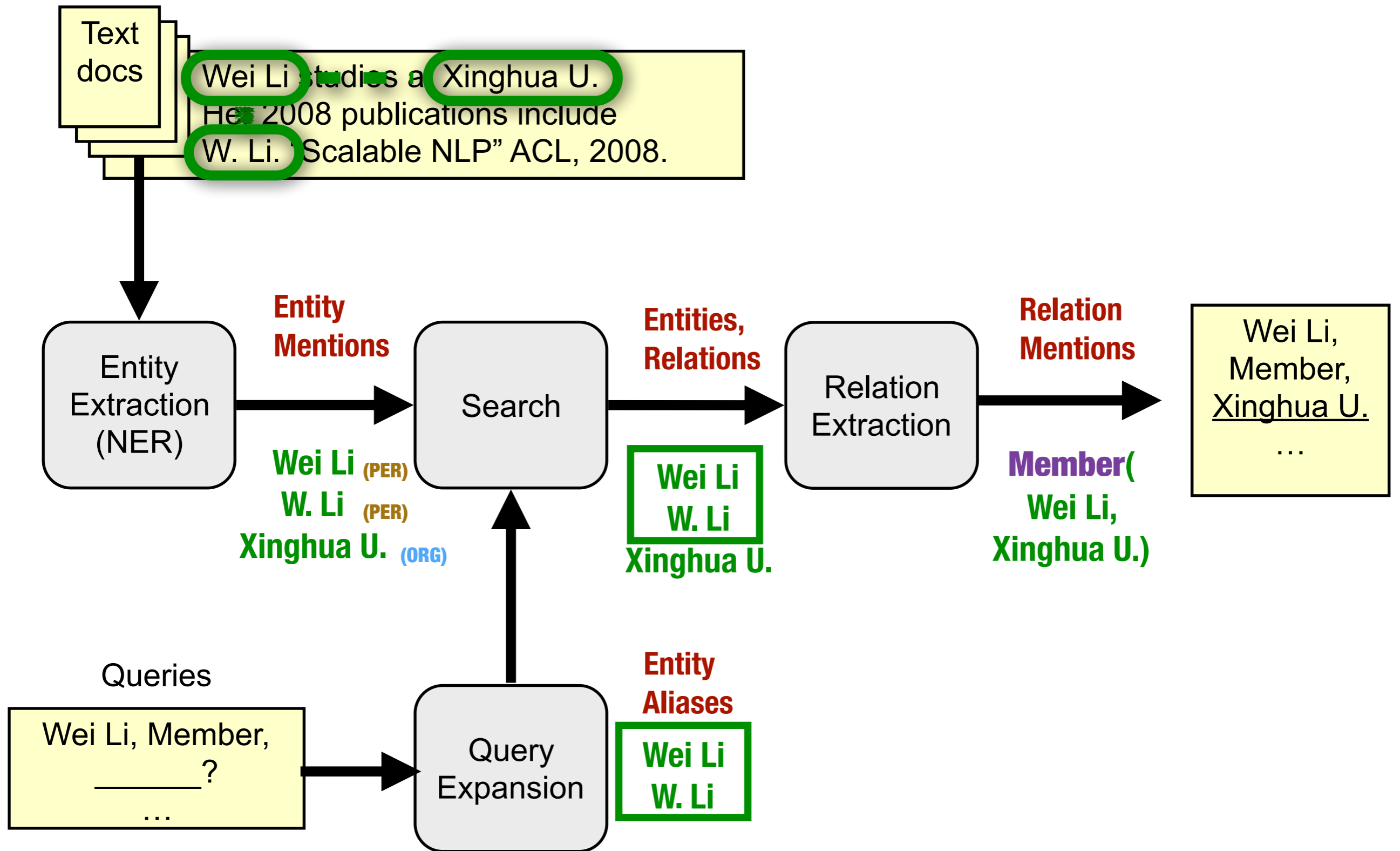# Multilingual Relation Extraction using Compositional Universal Schema

Haw-Shiuan Chang, Abdurrahman Munir, Ao Liu,
Johnny Tian-Zheng Wei, Aaron Traylor, Ajay Nagesh,
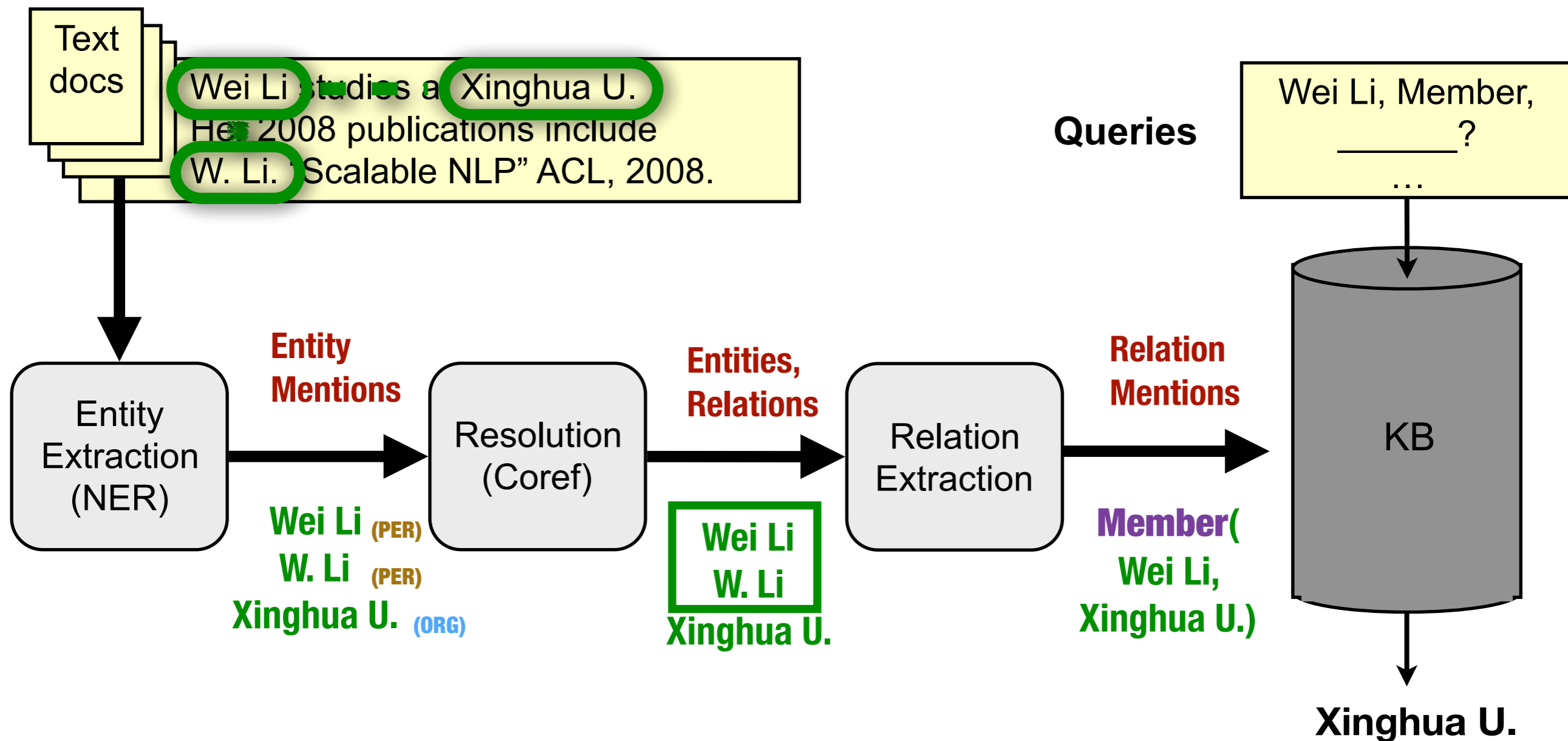Nicholas Monath, Patrick Verga, Emma Strubell, Andrew McCallum

UMassAmherst IESL

(Team ID: UMass_IESL)
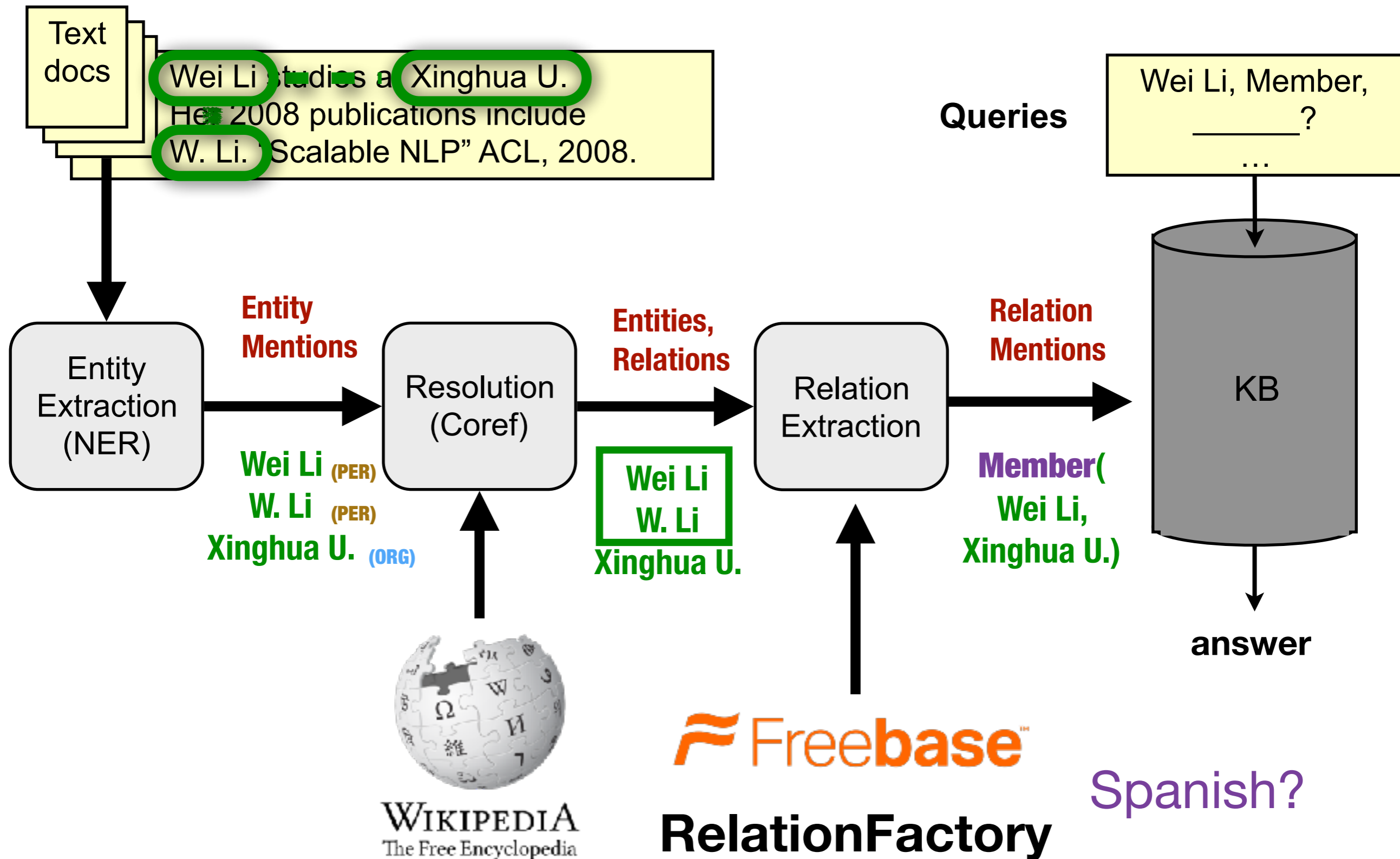
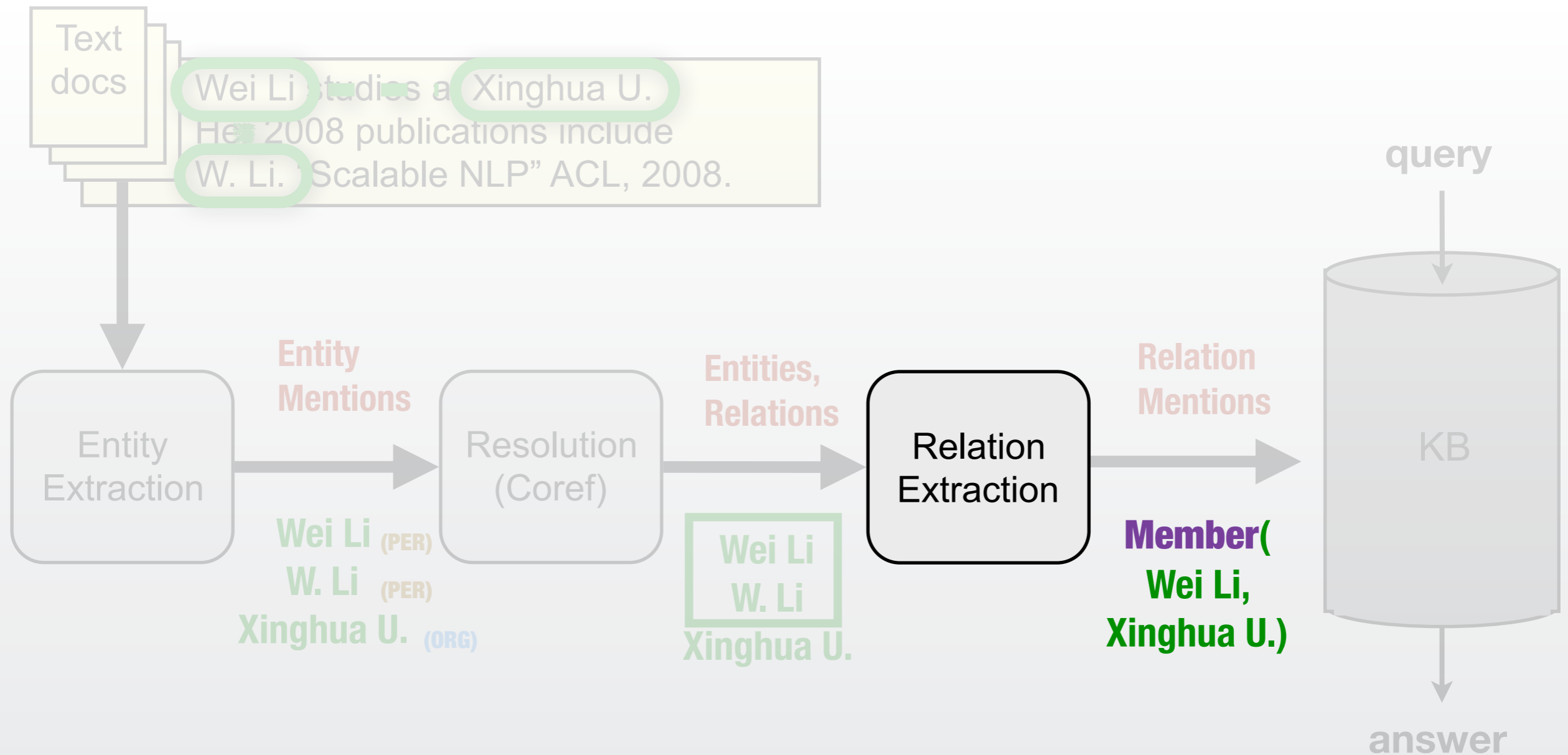These slides are mostly prepared by Patrick Verga

Slot Filling (SF)

Text docs

Wei Li studies at Xinghua U. Her 2008 publications include W. Li. "Scalable NLP" ACL, 2008.

Entity Extraction (NER)

Entity Mentions

Wei Li (PER)
W. Li (PER)
Xinghua U. (ORG)

Search

Entities, Relations

Wei Li
W. Li
Xinghua U.

Relation Extraction

Relation Mentions

Member(
Wei Li,
Xinghua U.)

Wei Li, Member, Xinghua U. …

Queries

Wei Li, Member, _____? …

Query Expansion

Entity Aliases

Wei Li
W. Li

Knowledge Base (KB) Construction

Text docs

Wei Li studies at Xinghua U. Her 2008 publications include W. Li. "Scalable NLP" ACL, 2008.

Queries

Wei Li, Member, _____? …

Entity Extraction (NER)

Entity Mentions

Wei Li (PER)
W. Li (PER)
Xinghua U. (ORG)

Resolution (Coref)

Entities, Relations

Wei Li
W. Li
Xinghua U.

Relation Extraction

Relation Mentions

Member(
Wei Li,
Xinghua U.)

KB

Xinghua U.

Resources

Relation Extraction

Text docs

Wei Li studies at Xinghua U.
Her 2008 publications include
W. Li. "Scalable NLP" ACL, 2008.

Entity Extraction

Entity Mentions

Resolution (Coref)

Entities, Relations

Relation Extraction

Relation Mentions

KB

query

answer

Wei Li (PER)
W. Li (PER)
Xinghua U. (ORG)

**Wei Li
W. Li
Xinghua U.**

**Member(
Wei Li,
Xinghua U.)**

Compositional Universal Schema

Verga, P., Belanger, D., Strubell, E., Roth, B., and McCallum, A. (2016). Multilingual relation extraction using compositional universal schema. In Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).

Universal Schema
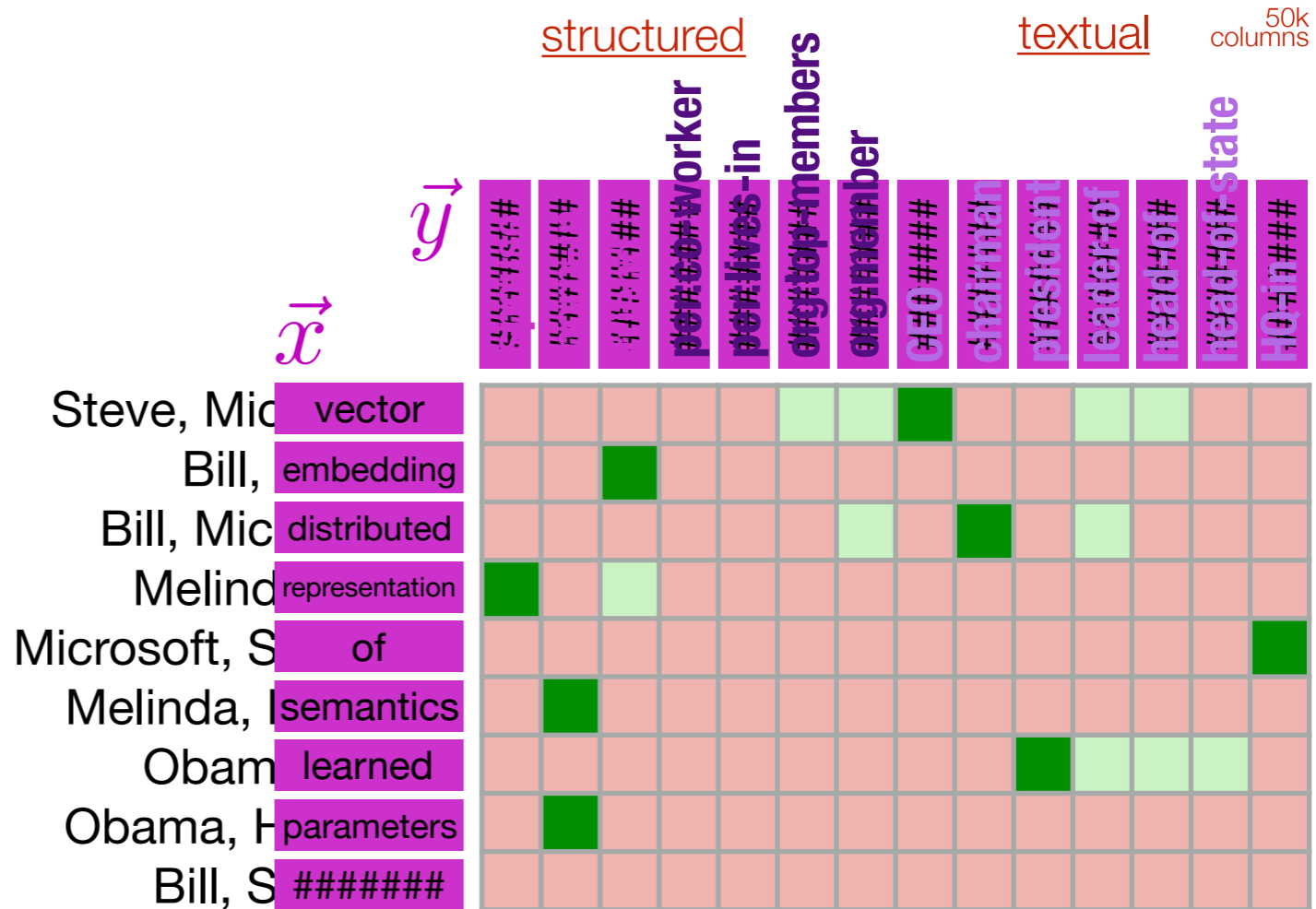
January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of **Microsoft**, the **Seattle**-headquartered software giant. His long-time friend, **Steve Balmer**, will take over as CEO of **Microsoft**. **Gates** will now focus on the charitable foundation he runs with **his** wife **Melinda French Gates**. **Bill** and **Melinda** were married in a ceremony in **Hawaii**, rather than her hometown of **Dallas**. **Gates** moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside **Seattle**.
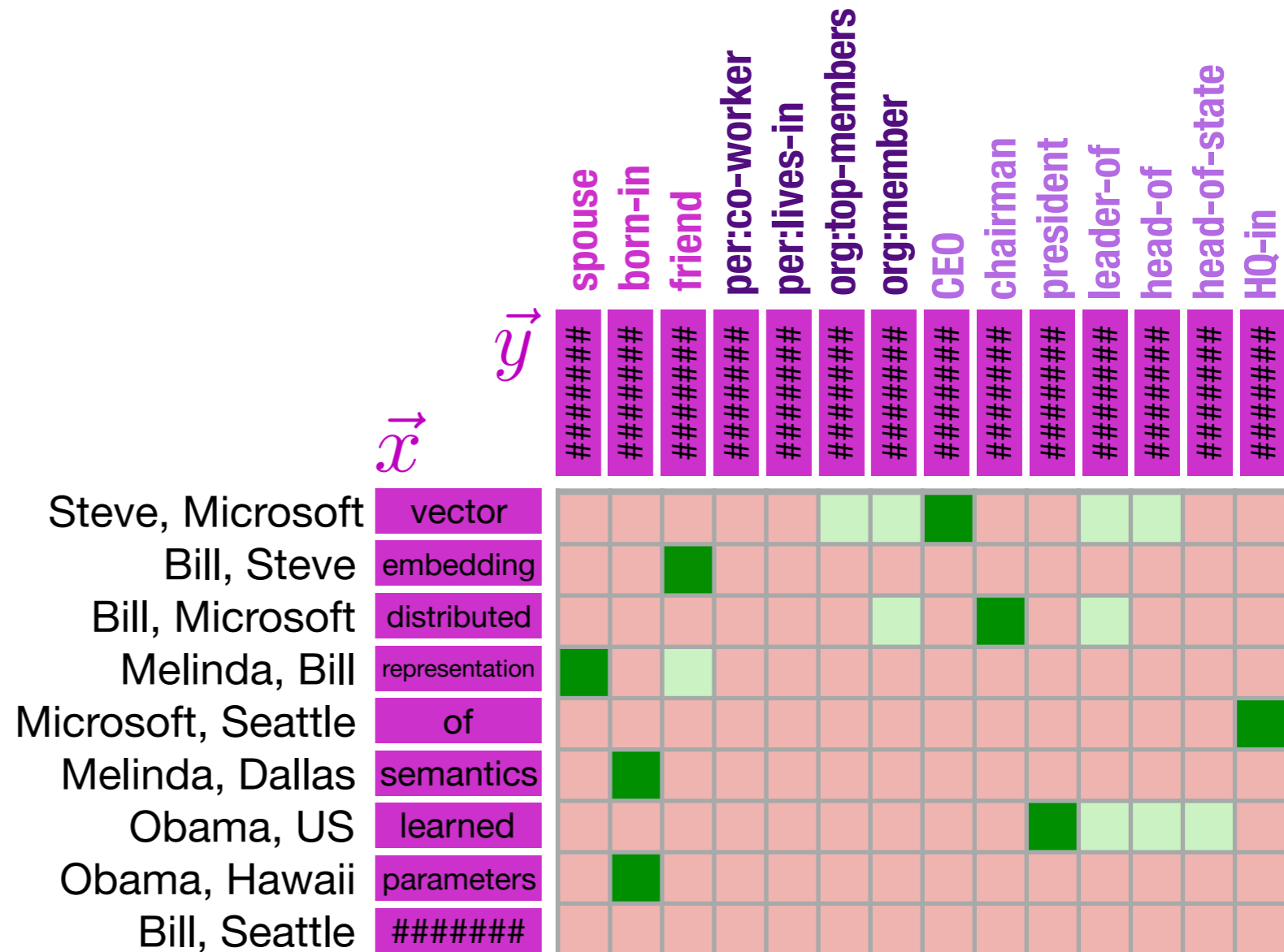
Relation Types

structured   textual   50k columns

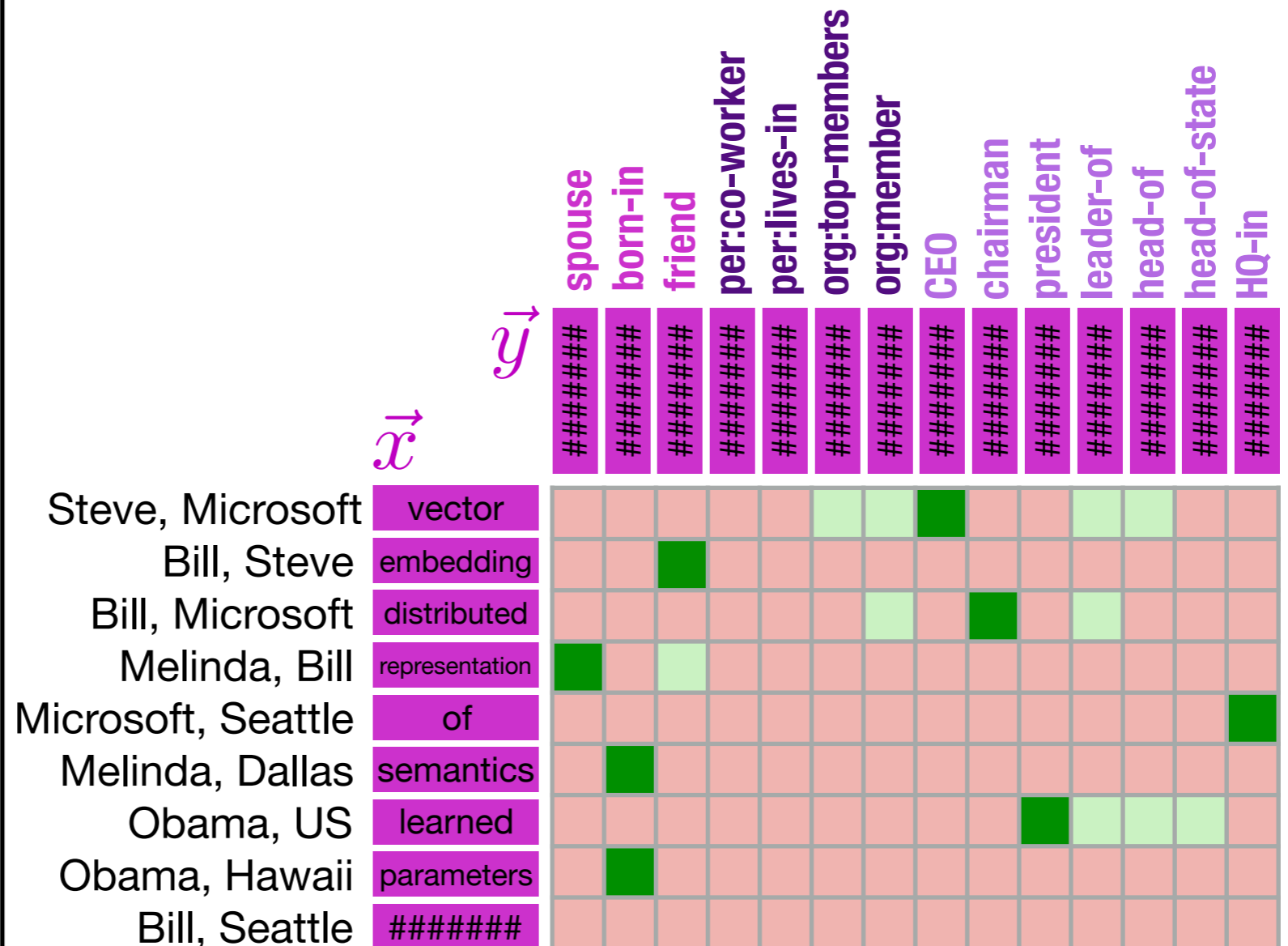|	spouse	born-in	friend	per:co-worker	per:lives-in	org:top-members	org:member	CEO	chairman	president	leader-of	head-of	head-of-state	HQ-in
Steve, Microsoft						▨	▨	▩			▨	▨		
Bill, Steve			▩											
Bill, Microsoft							▨			▩		▨		
Melinda, Bill	▩		▨											▩
Microsoft, Seattle														▩
Melinda, Dallas		▩												
Obama, US										▩	▨	▨		
Obama, Hawaii		▩												
Bill, Seattle														

January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of **Microsoft**, the **Seattle**-headquartered software giant. His long-time friend, **Steve Balmer**, will take over as CEO of **Microsoft**. **Gates** will now focus on the charitable foundation he runs with **his** wife **Melinda French Gates**. **Bill** and **Melinda** were married in a ceremony in **Hawaii**, rather than her hometown of **Dallas**. **Gates** moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside **Seattle**.

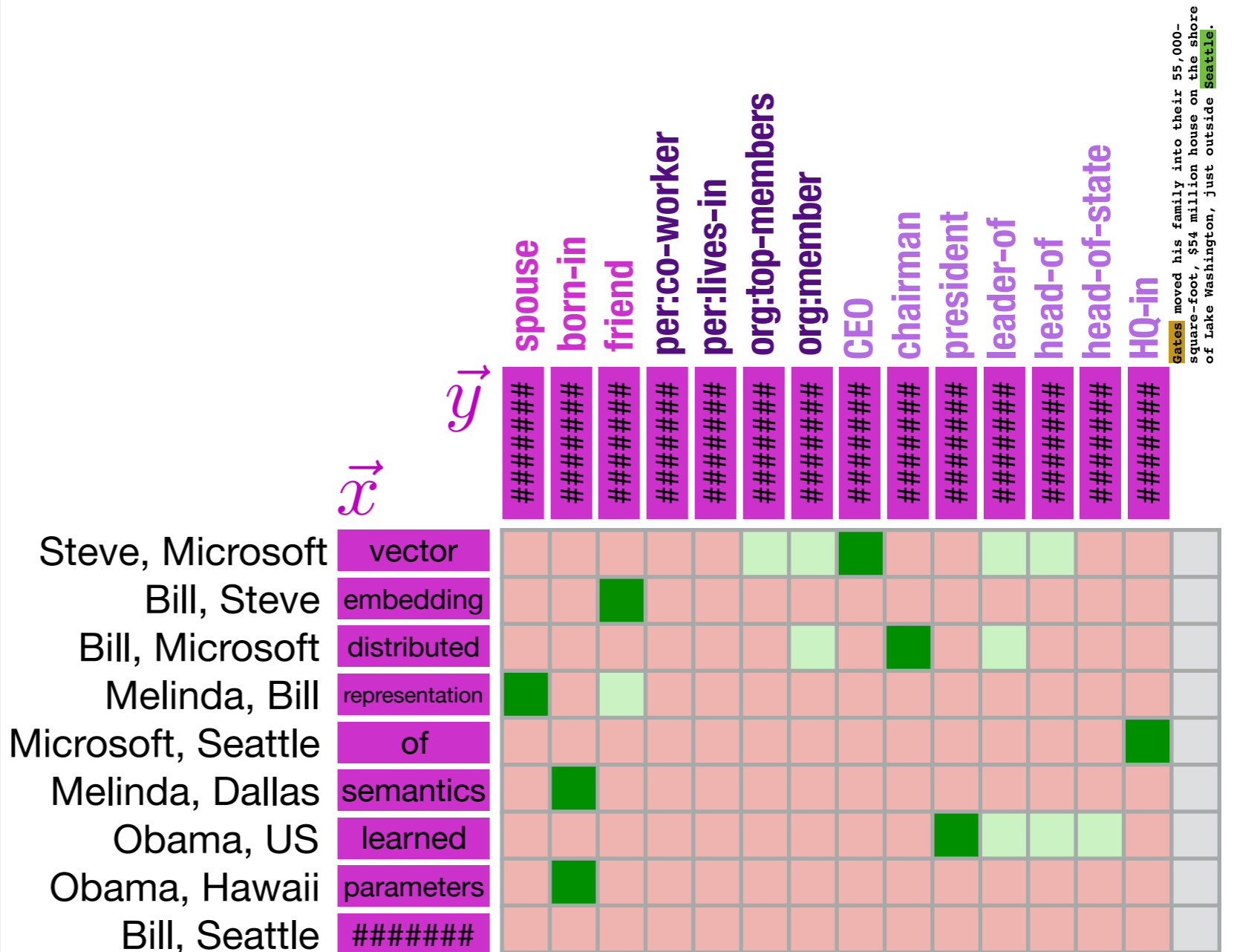$$\mathbb{P}\left((s, r, o)\right) = \sigma\left(u_{s,o}^{\top} v_r\right)$$
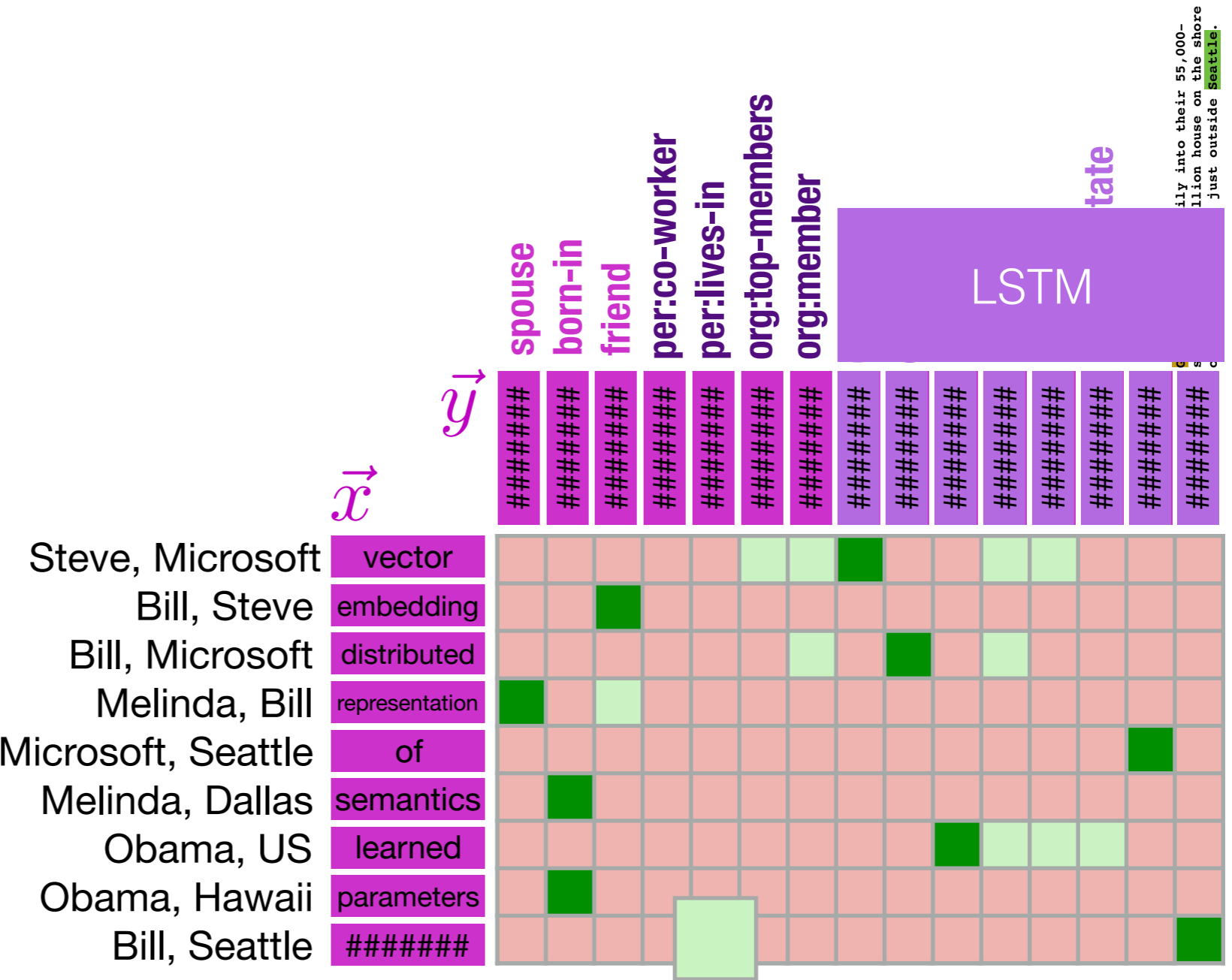
8

January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of **Microsoft**, the **Seattle**-headquartered software giant. His long-time friend, **Steve Balmer**, will take over as CEO of **Microsoft**. **Gates** will now focus on the charitable foundation he runs with **his** wife **Melinda French Gates**. **Bill** and **Melinda** were married in a ceremony in **Hawaii**, rather than her hometown of **Dallas**. **Gates** moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside **Seattle**.

$\vec{y}$

	spouse	born-in	friend	per:co-worker	per:lives-in	org:top-members	org:member	CEO	chairman	president	leader-of	head-of	head-of-state	HQ-in
Steve, Microsoft — vector								■						
Bill, Steve — embedding			■											
Bill, Microsoft — distributed										■				
Melinda, Bill — representation	■													
Microsoft, Seattle — of														■
Melinda, Dallas — semantics		■												
Obama, US — learned											■			
Obama, Hawaii — parameters		■												
Bill, Seattle — #######														

$\vec{x}$

9

Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside **Seattle**.
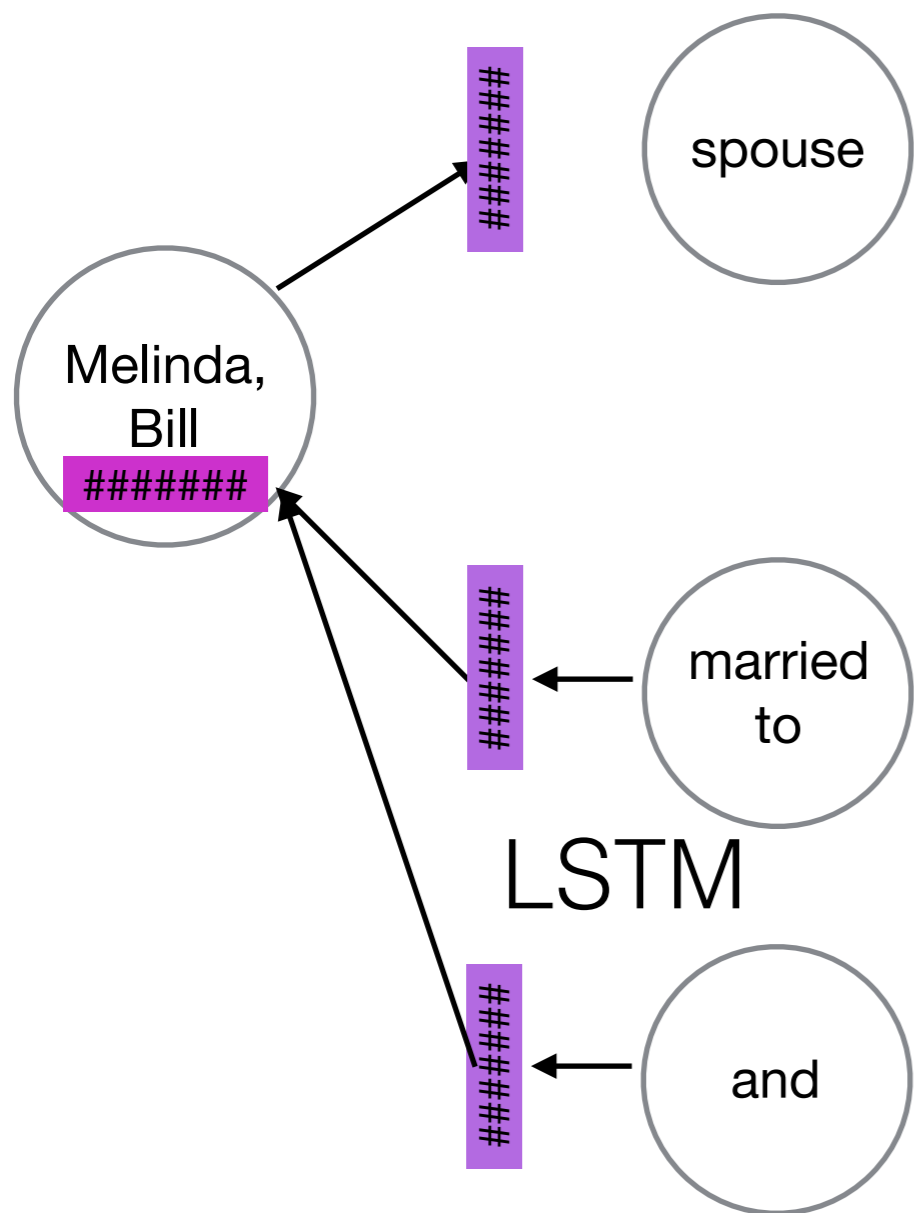
January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of **Microsoft**, the **Seattle**-headquartered software giant. His long-time friend, **Steve Balmer**, will take over as CEO of **Microsoft**. **Gates** will now focus on the charitable foundation he runs with **his** wife **Melinda French Gates**. **Bill** and **Melinda** were married in a ceremony in **Hawaii**, rather than her hometown of **Dallas**. **Gates** moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside **Seattle**.

$\vec{y}$

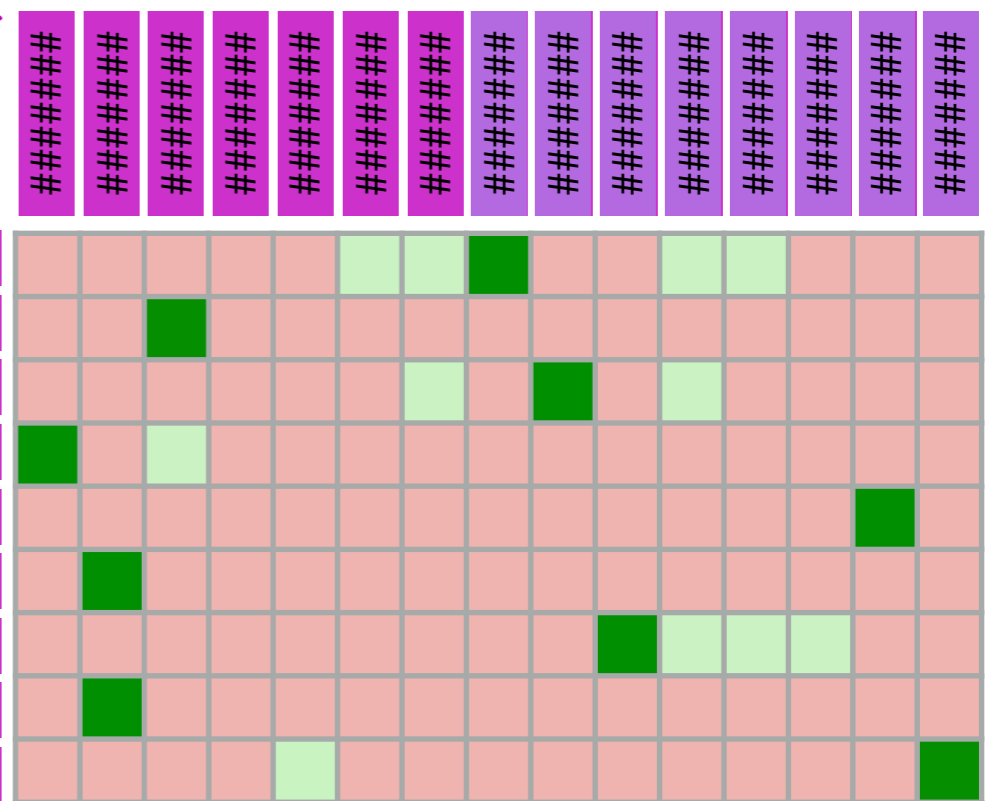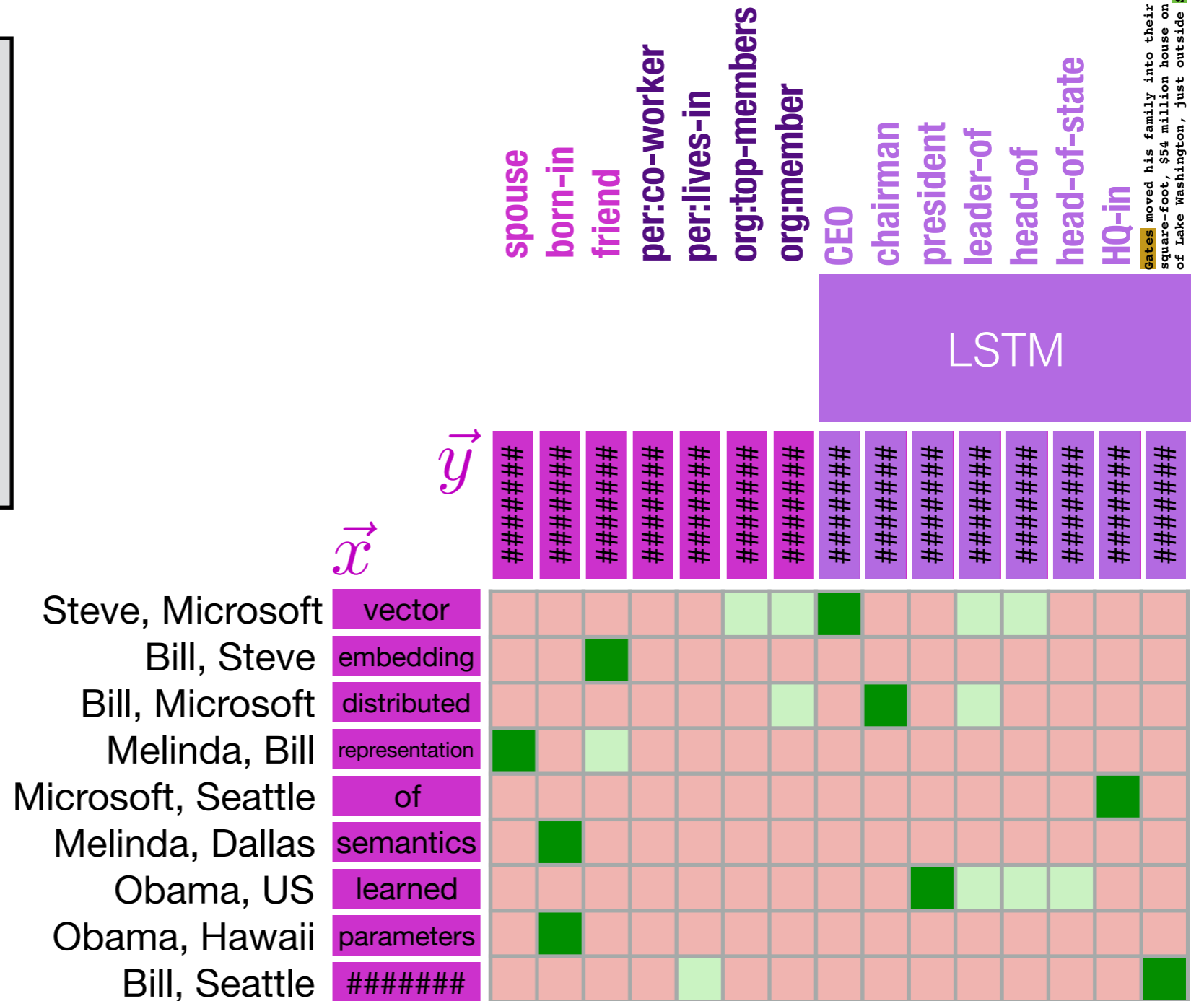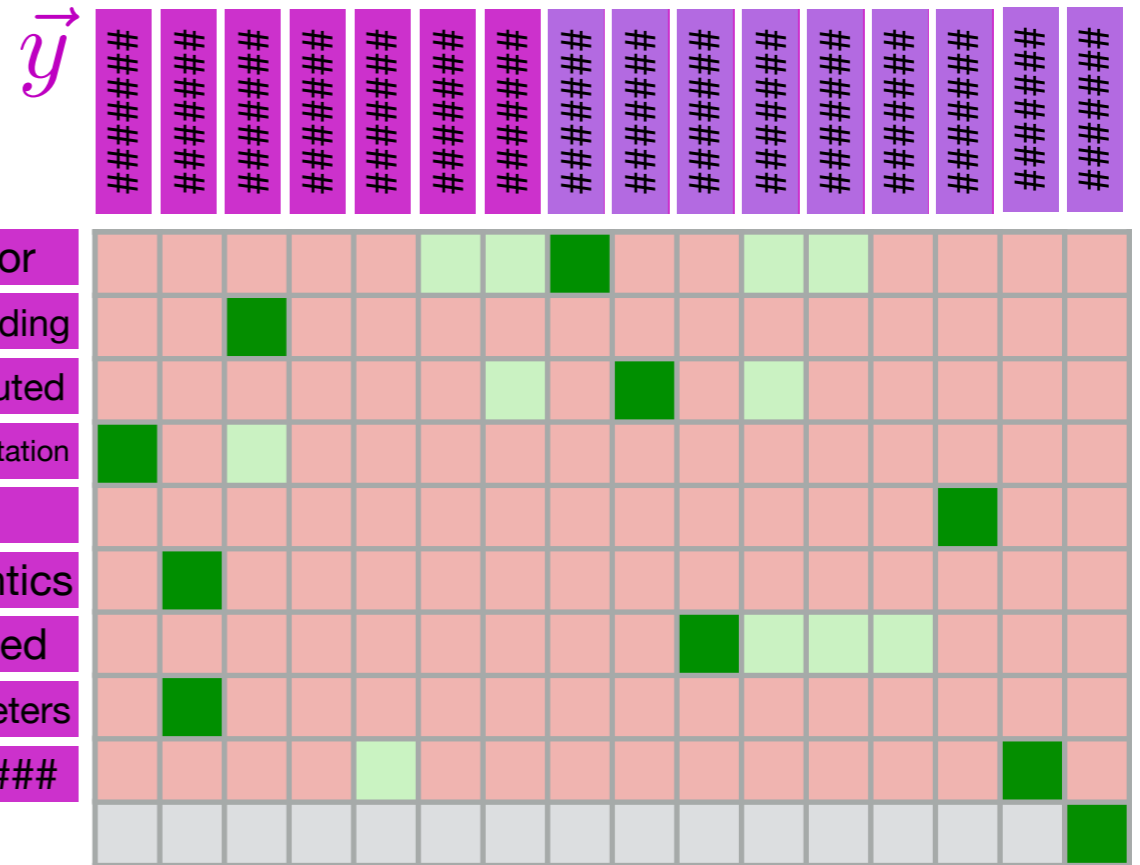| $\vec{x}$		spouse	born-in	friend	per:co-worker	per:lives-in	org:top-members	org:member	CEO	chairman	president	leader-of	head-of	head-of-state	HQ-in
Steve, Microsoft	vector						light	light	dark			light			
Bill, Steve	embedding			dark											
Bill, Microsoft	distributed					light				dark					
Melinda, Bill	representation	dark		light											
Microsoft, Seattle	of														dark
Melinda, Dallas	semantics		dark												
Obama, US	learned											dark	light	light	
Obama, Hawaii	parameters		dark												
Bill, Seattle	#######														

10

January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.

$\vec{y}$

$\vec{x}$		spouse	born-in	friend	per:co-worker	per:lives-in	org:top-members	org:member	CEO	chairman	president	leader-of	head-of	head-of-state	HQ-in
Steve, Microsoft	vector						●	●	■			●	●		
Bill, Steve	embedding			■											
Bill, Microsoft	distributed								●	■		●			
Melinda, Bill	representation	■		●											
Microsoft, Seattle	of														■
Melinda, Dallas	semantics		■												
Obama, US	learned										■	●		●	
Obama, Hawaii	parameters		■												
Bill, Seattle	#######														

11

January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of **Microsoft**, the **Seattle**-headquartered software giant. His long-time friend, **Steve Balmer**, will take over as CEO of **Microsoft**. **Gates** will now focus on the charitable foundation he runs with **his** wife **Melinda French Gates**. **Bill** and **Melinda** were married in a ceremony in **Hawaii**, rather than her hometown of **Dallas**. **Gates** moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside **Seattle**.

$\vec{y}$

	spouse	born-in	friend	per:co-worker	per:lives-in	org:top-members	org:member								LSTM
$\vec{x}$	######	######	######	######	######	######	######	###	###	###	###	###	###	###	###
Steve, Microsoft — vector					■	■	■			■					
Bill, Steve — embedding			■												
Bill, Microsoft — distributed							■		■						
Melinda, Bill — representation	■		■												
Microsoft, Seattle — of													■		
Melinda, Dallas — semantics		■													
Obama, US — learned										■	■		■		
Obama, Hawaii — parameters		■													
Bill, Seattle — #######						■									■

12

13

query: **spouse**

Joe NewEntity worked with his wife Jane NewEntity on relation extraction.

spouse
born-in
friend
per:co-worker
per:lives-in
org:top-members
org:member
CEO
chairman
president
leader-of
head-of
head-of-state
HQ-in

Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.

LSTM

$\vec{y}$

$\vec{x}$		spouse	born-in	friend	per:co-worker	per:lives-in	org:top-members	org:member	CEO	chairman	president	leader-of	head-of	head-of-state	HQ-in
Steve, Microsoft	vector														
Bill, Steve	embedding														
Bill, Microsoft	distributed														
Melinda, Bill	representation														
Microsoft, Seattle	of														
Melinda, Dallas	semantics														
Obama, US	learned														
Obama, Hawaii	parameters														
Bill, Seattle	#######														

14

query: **spouse**

Joe NewEntity worked with his wife Jane NewEntity on relation extraction.

spouse
born-in
friend
per:co-worker
per:lives-in
org:top-members
org:member
CEO
chairman
president
leader-of
head-of
head-of-state
HQ-in

Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.

Joe NewEntity worked with his wife Jane NewEntity on relation extraction.

LSTM

$\vec{y}$

$\vec{x}$

		spouse	born-in	friend	per:co-worker	per:lives-in	org:top-members	org:member	CEO	chairman	president	leader-of	head-of	head-of-state	HQ-in		
Steve, Microsoft	vector						░	░	■			░	░				
Bill, Steve	embedding			■													
Bill, Microsoft	distributed						░		■		░						
Melinda, Bill	representation	■		░													
Microsoft, Seattle	of														■		
Melinda, Dallas	semantics		■														
Obama, US	learned									■	░	░		░			
Obama, Hawaii	parameters		■														
Bill, Seattle	#######				░										■		
																■	

15

query: **spouse**

Joe NewEntity worked with his wife Jane NewEntity on relation extraction.

16

Spanish
Relation Extraction

Multilingual Relation Extraction

Multilingual Universal Schema

24

cat gato

married casado

spouse / esposa

presidente

resident

spouse born-in friend per:co-worker per:lives-in org:top-members org:member CEO chairman president leader-of head-of married-to HQ-in esposa presidente vive en residente fundador

LSTM

$\vec{y}$

$\vec{x}$

Steve, Microsoft — vector
Bill, Steve — embeddin
Bill, Microsoft — distribute
Melinda, Bill — #######
Microsoft, Seattle — of
Melinda, Dallas — semanti
Obama, US — learned
Obama, Michelle — #######
Bill, Seattle — #######
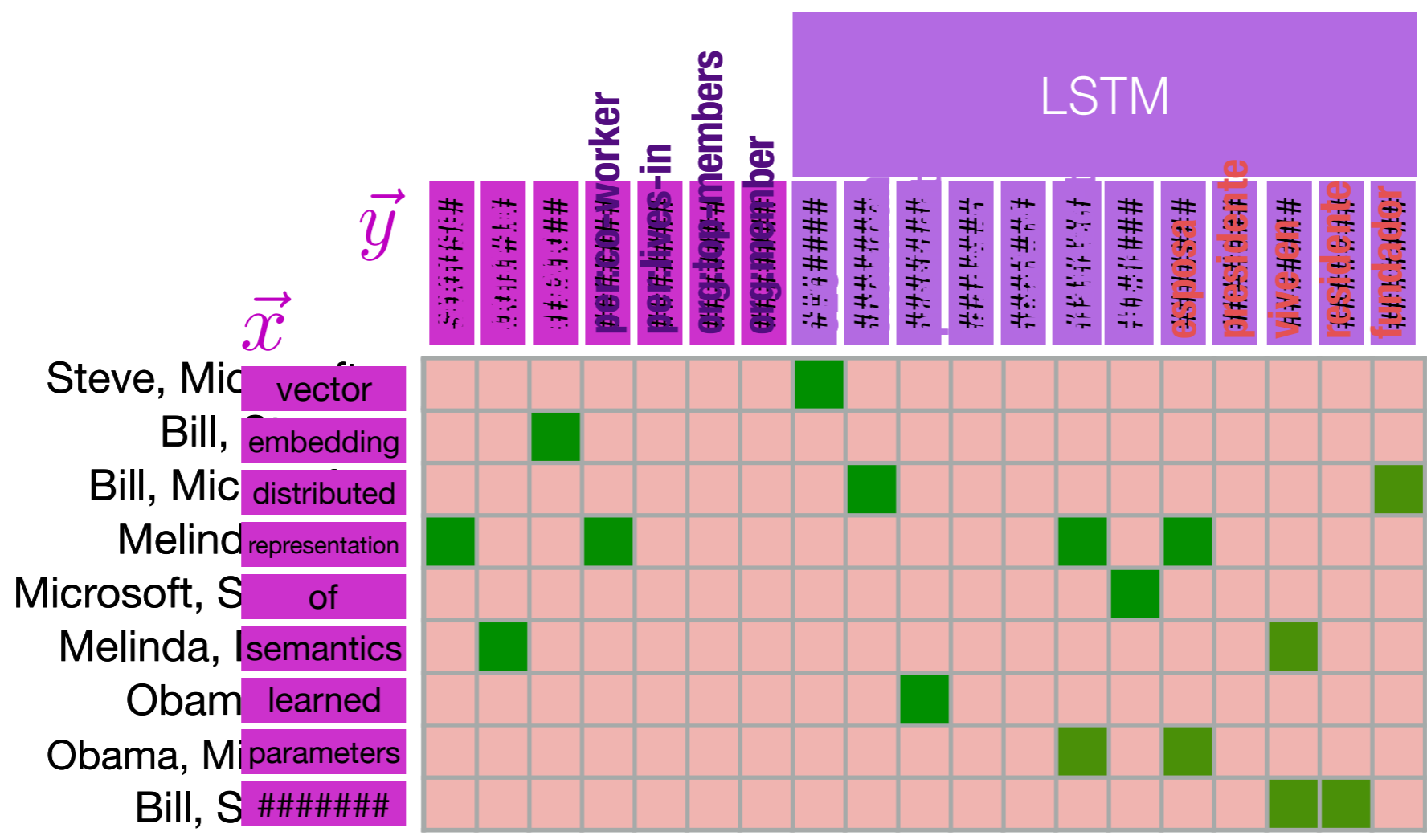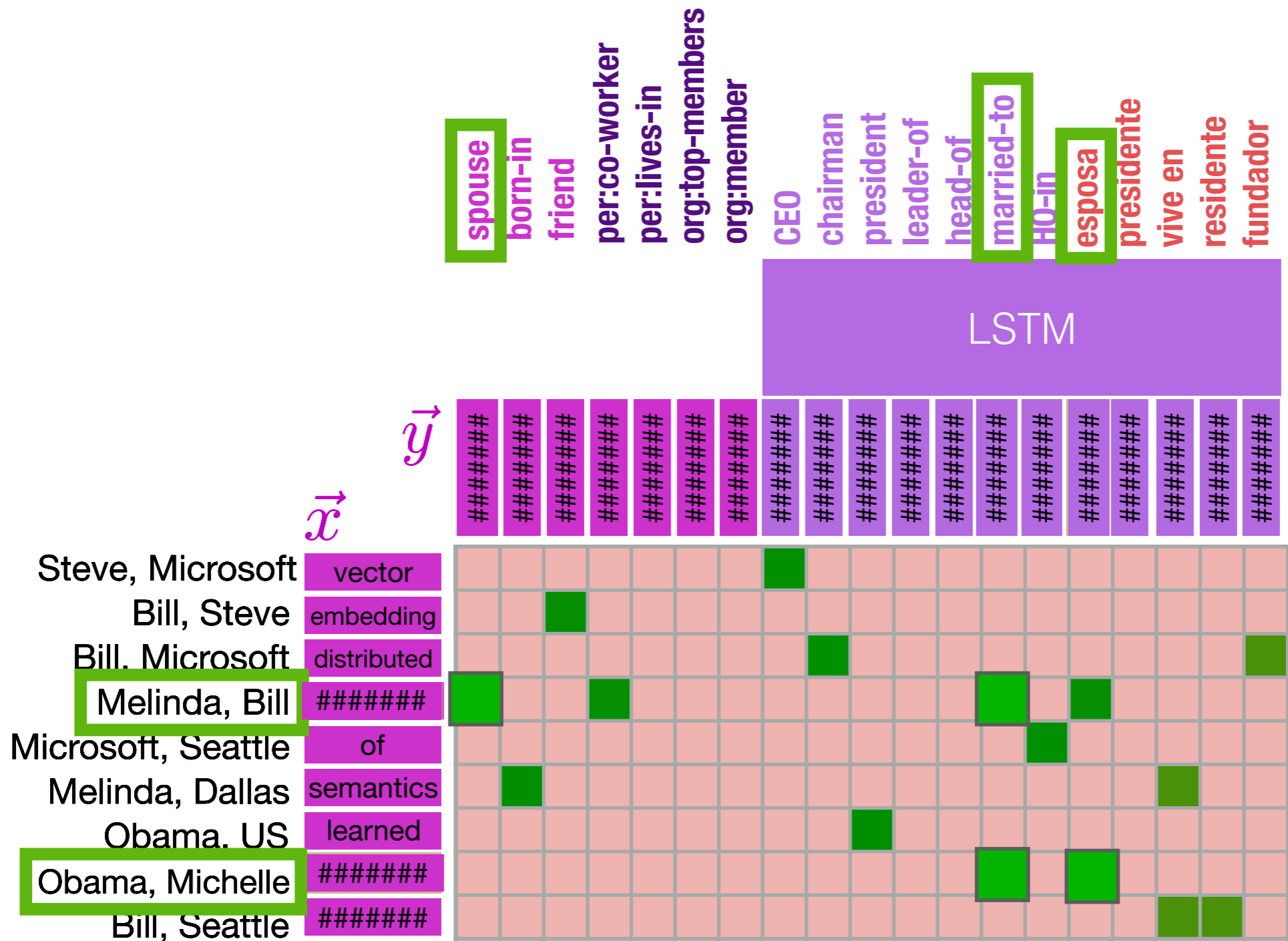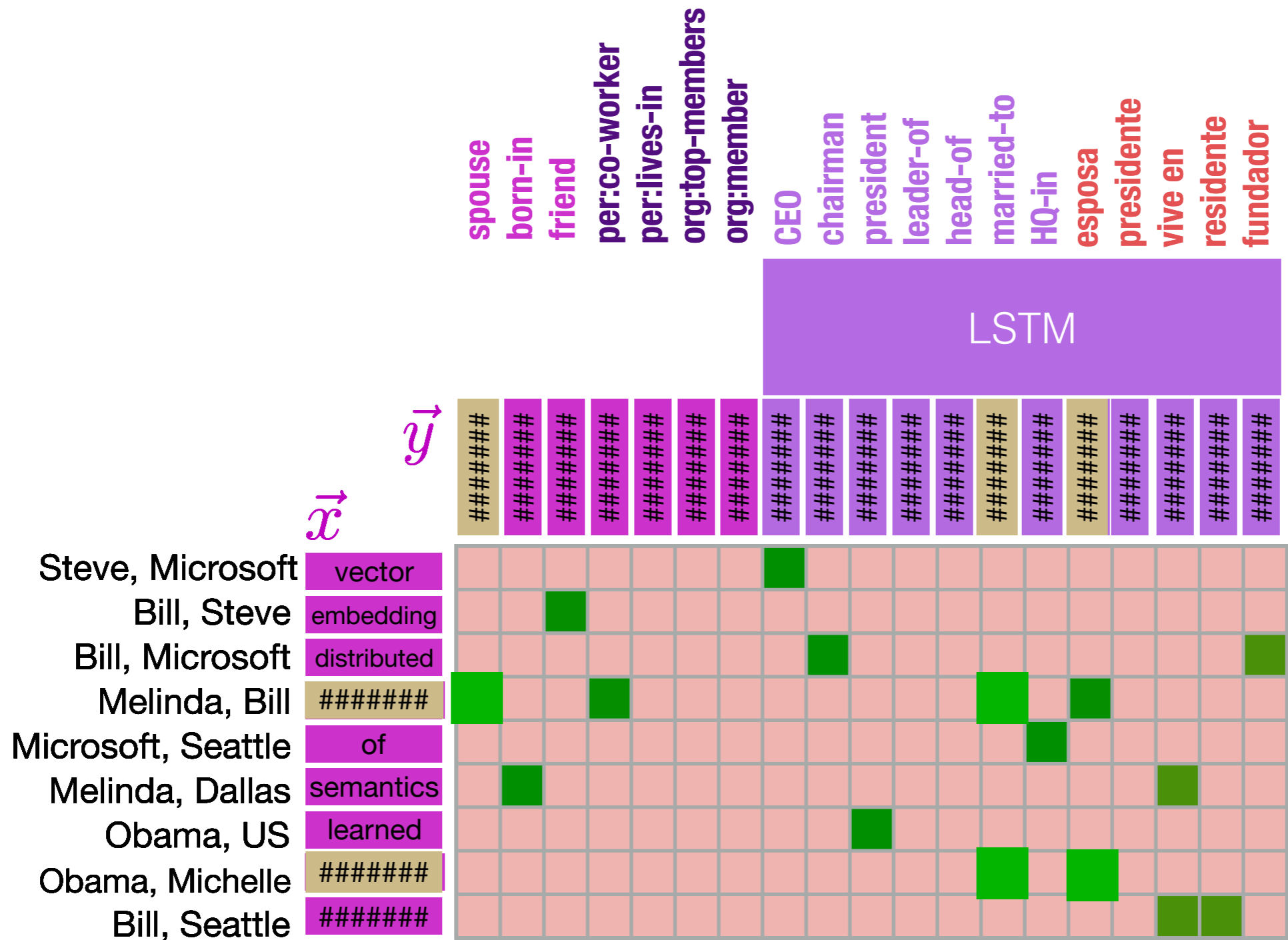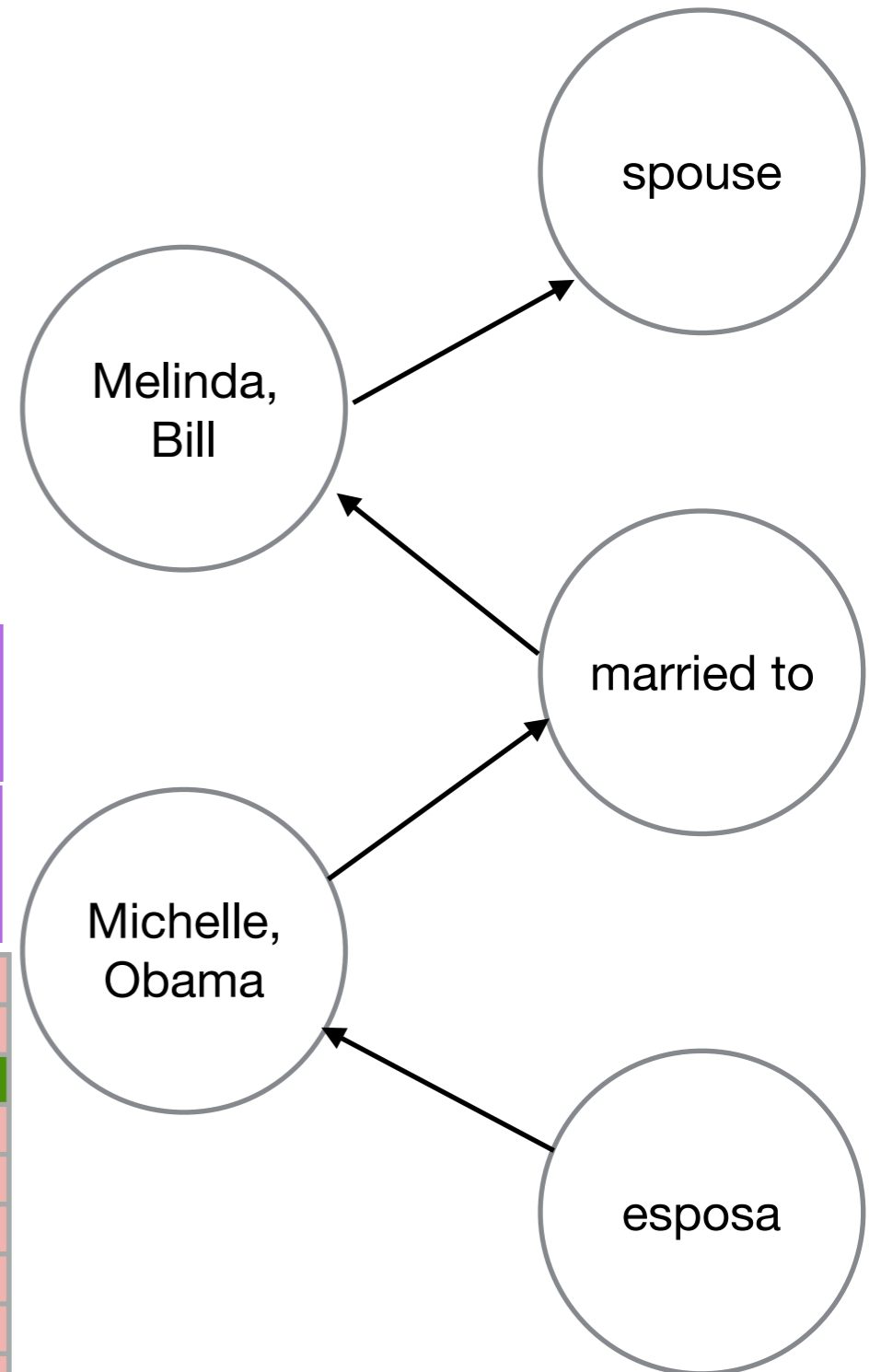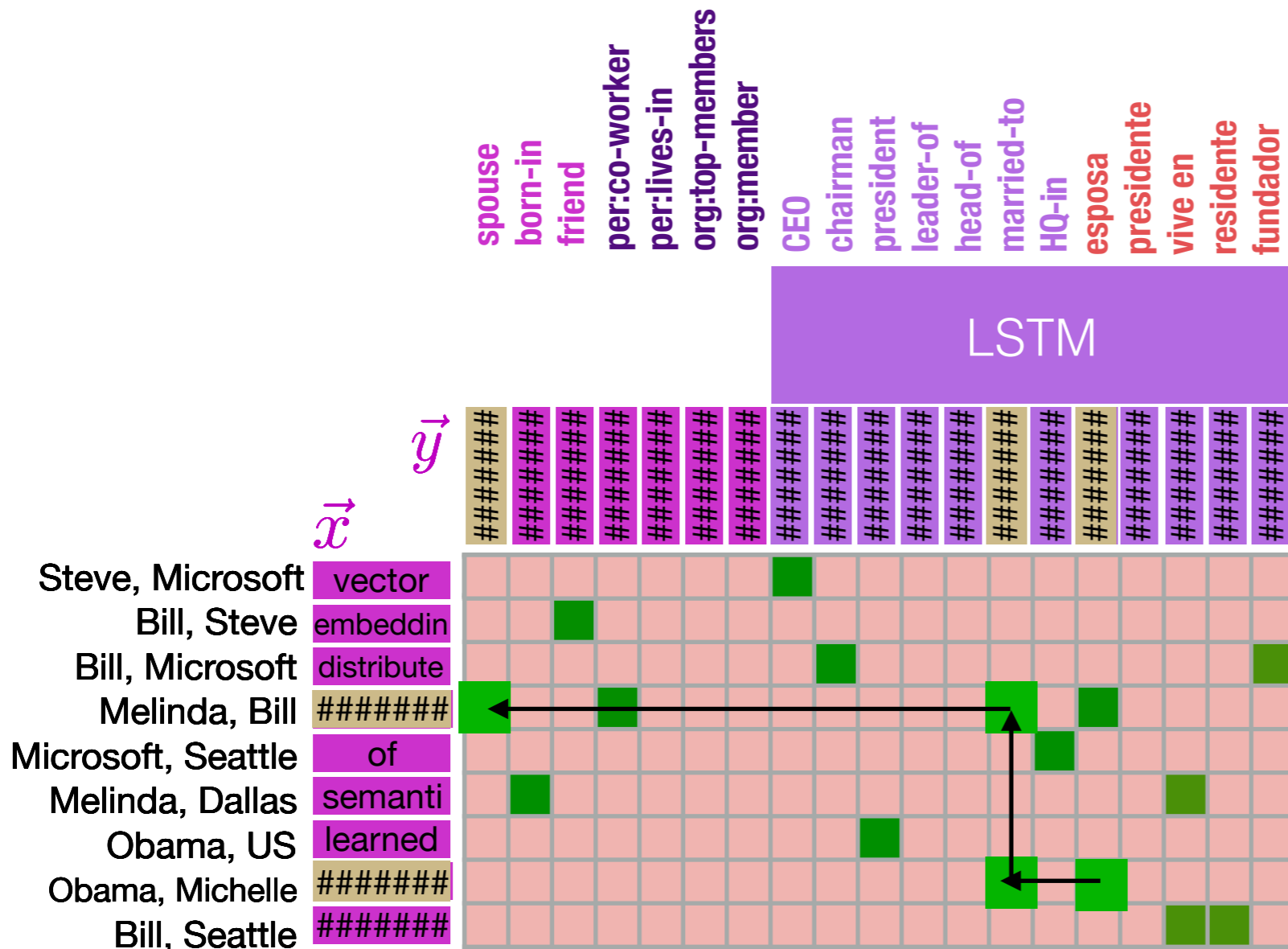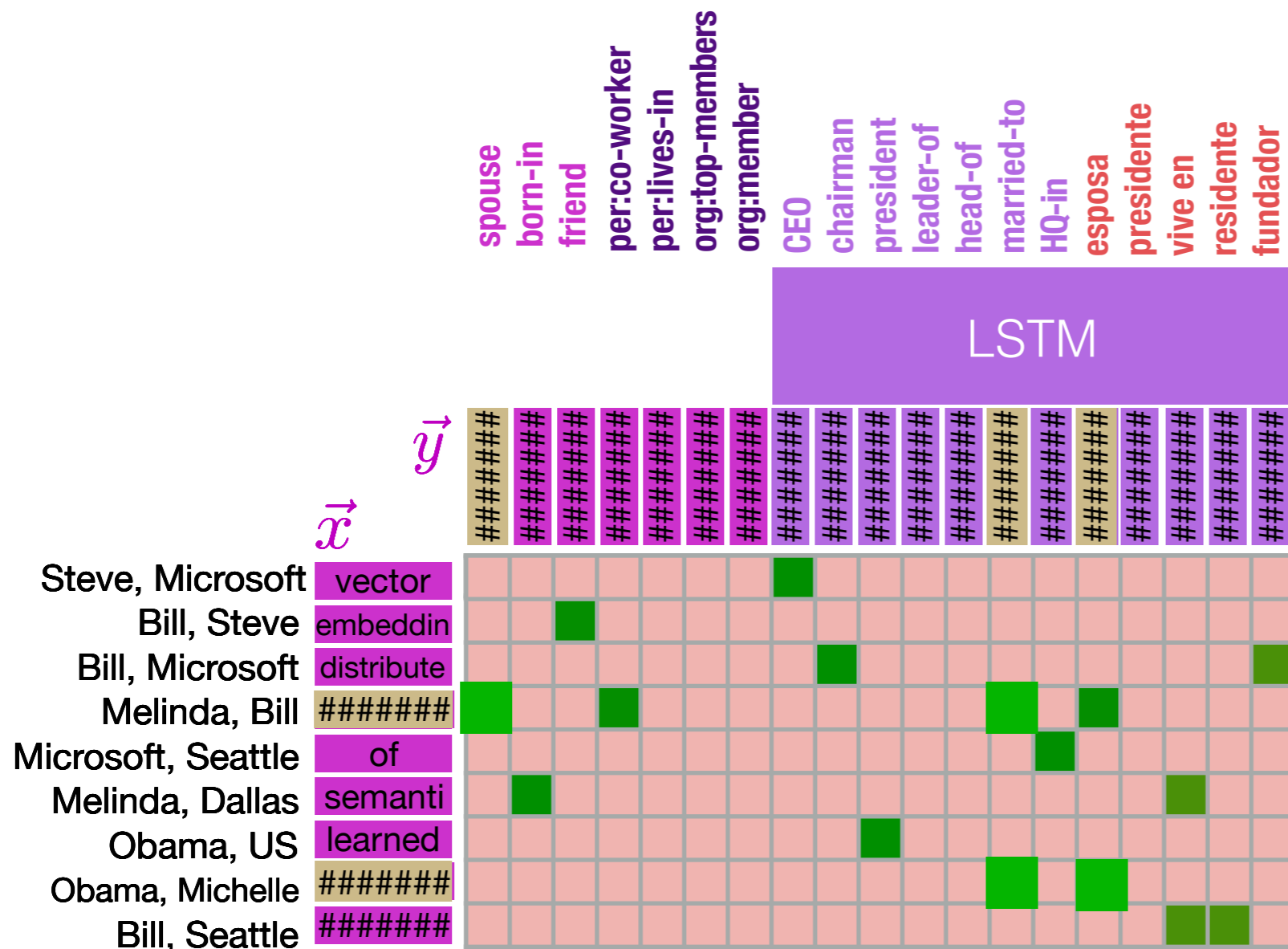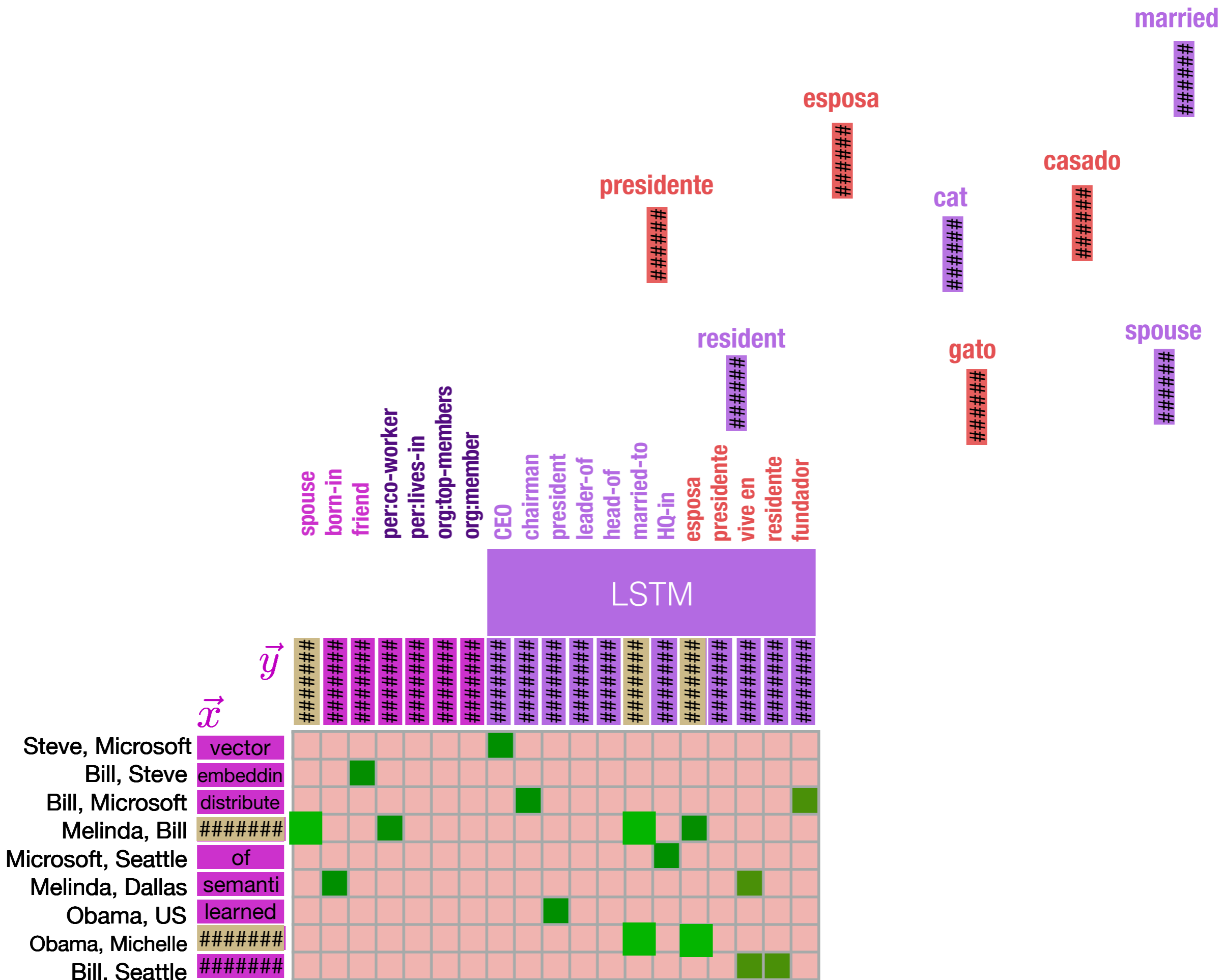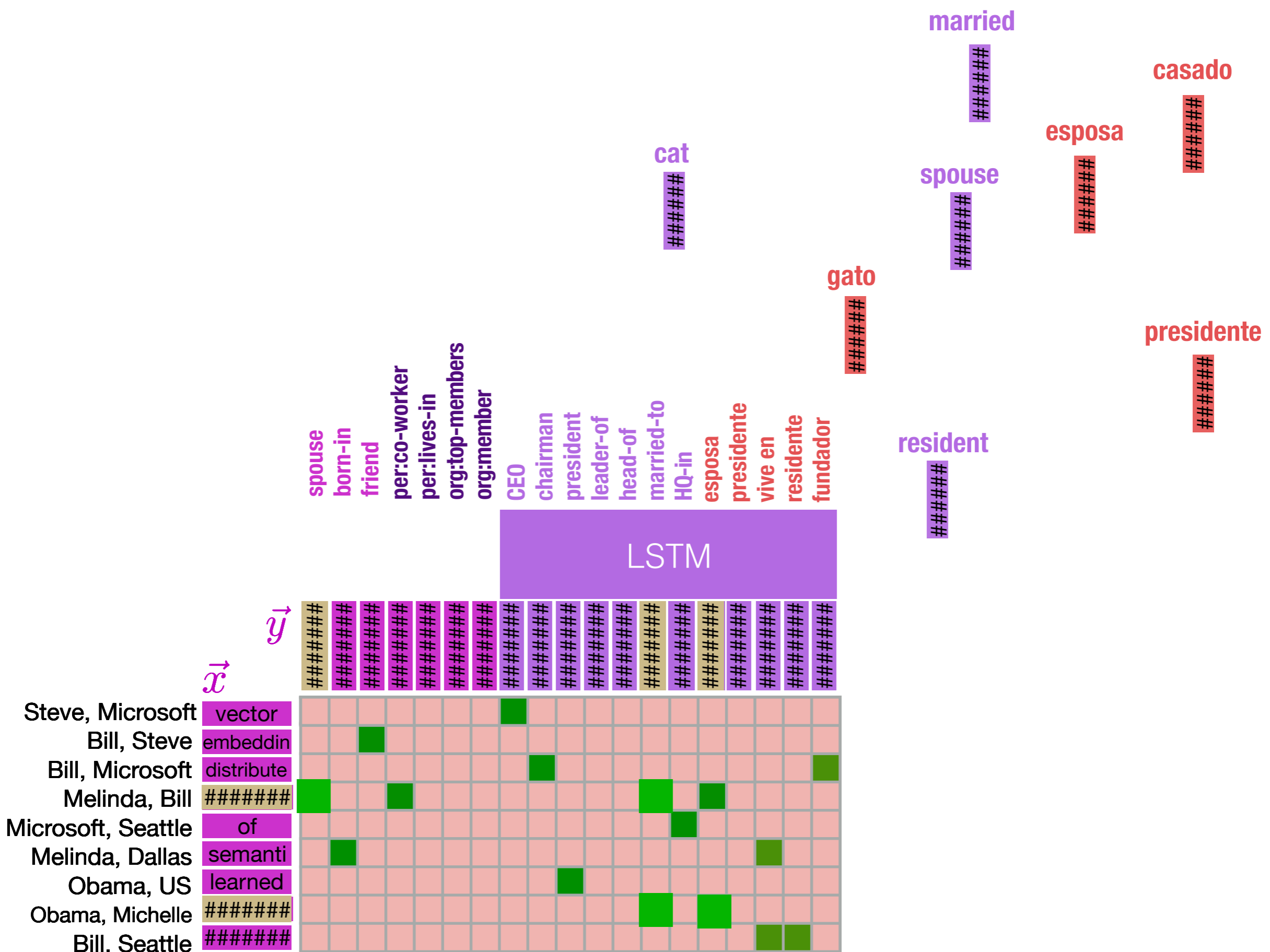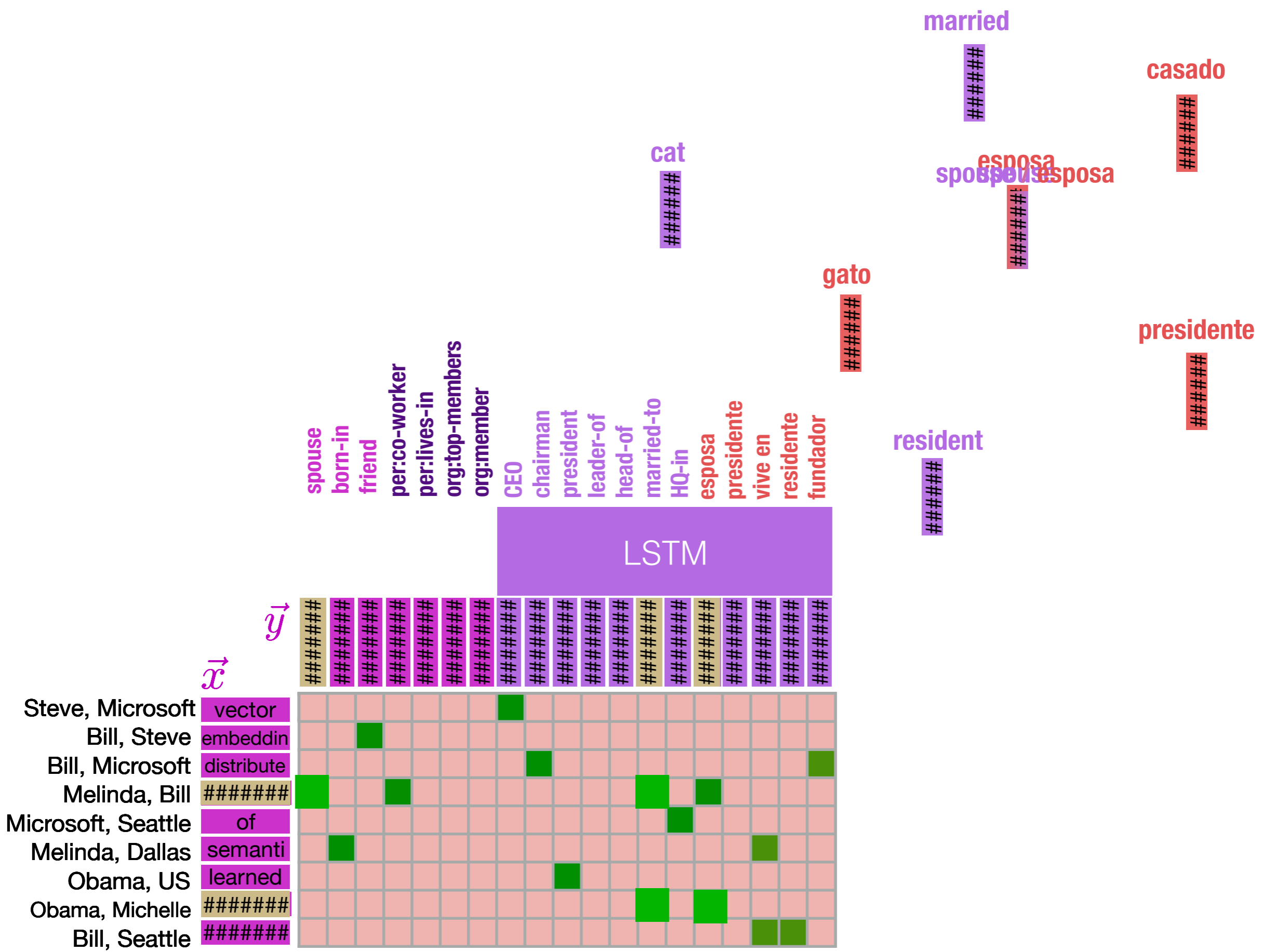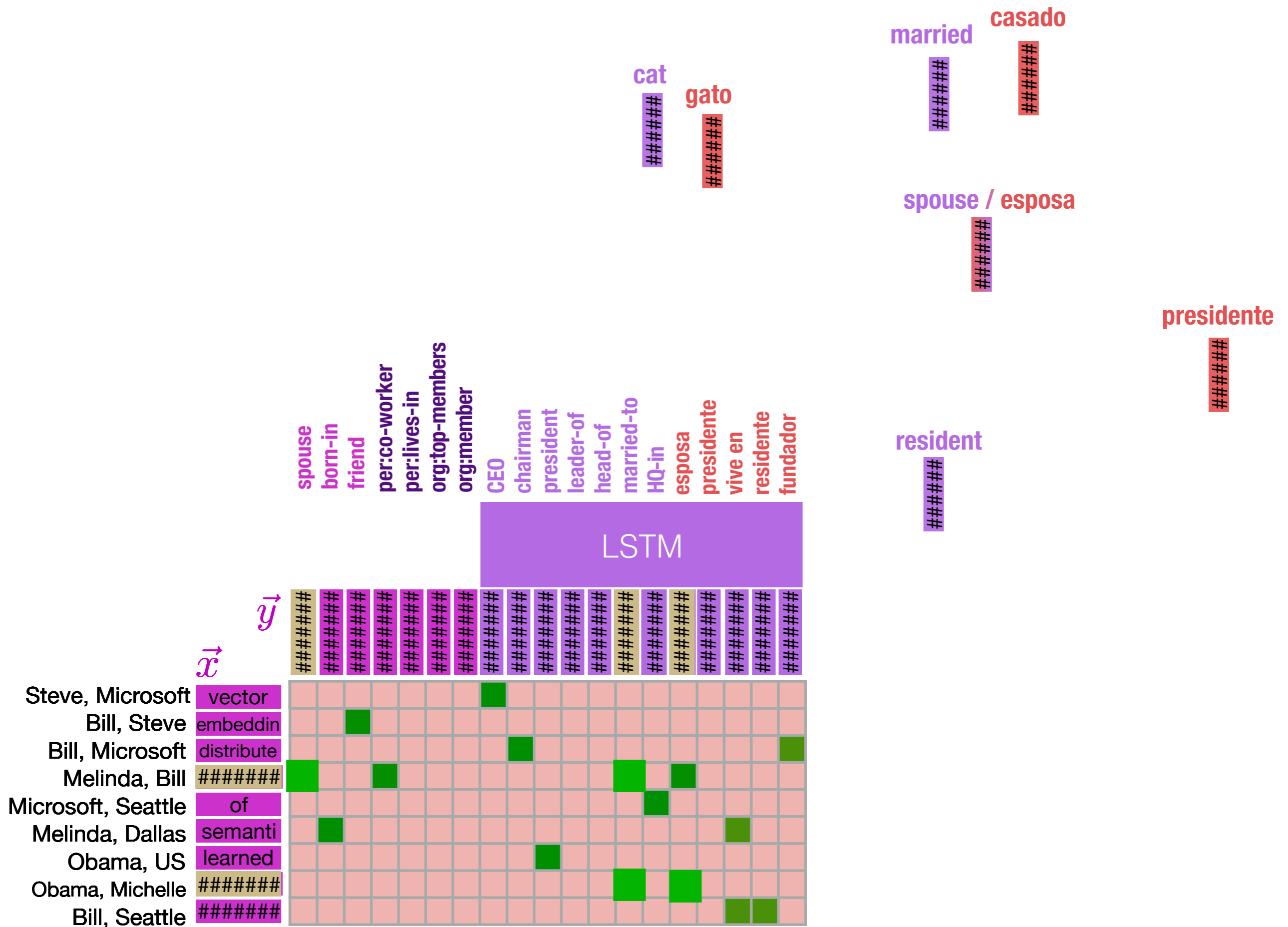
Search Engine Supervision

Search Engine as a Resource

Text docs

Wei Li studies at Xinghua U.
Her 2008 publications include
W. Li. "Scalable NLP" ACL, 2008.

query

Entity Mentions

Entities, Relations

Relation Mentions

Entity Extraction → Resolution (Coref) → Relation Extraction → KB

Wei Li (PER)
W. Li (PER)
Xinghua U. (ORG)

Wei Li
W. Li
Xinghua U.

Member(
Wei Li,
Xinghua U.)

Google

Freebase

RelationFactory

WIKIPEDIA
The Free Encyclopedia

answer

Distant Supervision

Limitations of Distant Supervision

| Barack Obama | Entity 1 | | Entity 2 | Michelle Obama |

wife ← High frequency signal

and ← Global background noise ← Not expressing a relation

president ← Entity background noise ← Freebase selection bias

lives with ← High correlation noise ← Only valid in some entity pairs

understands ← Other relation noise ← They might have other relations

married to ← Low frequency signal

Asking in another direction?

Barack Obama … wife … _____

Barack Obama … and … _____

P(___ = Michelle Obama) ↑

P(___ = Michelle Obama) ↓

Corpus

Key phrases
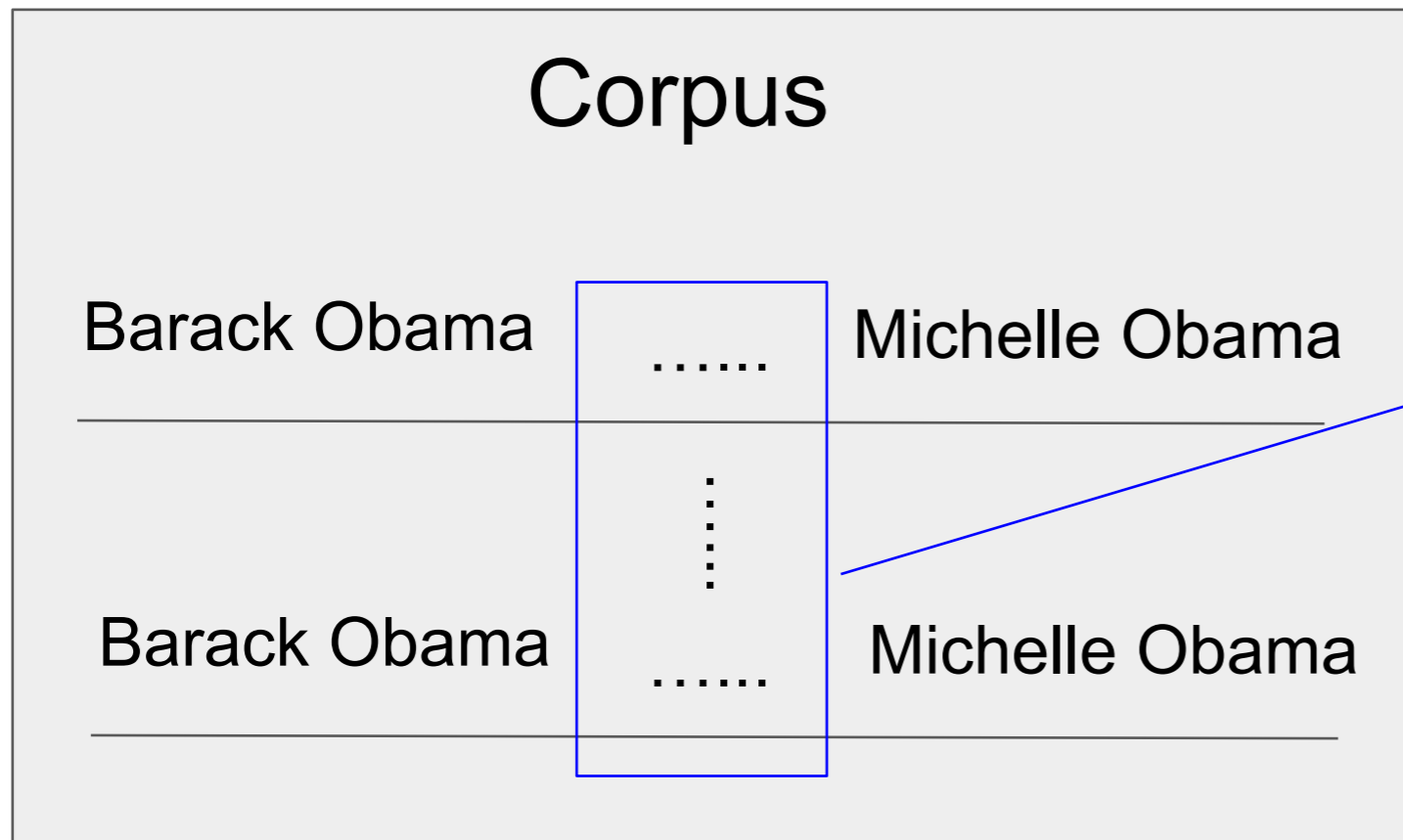
wife

and

president

lives with

understands

married to

Entity 1

Barack Obama

Entity 2

Michelle Obama?

Hard to accurately estimate this when corpus is not large enough

Asking Google!

Noise Filtering Examples

Distant supervision
(after applying several noises
removal techniques)

After applying search
engine supervision

wife
"his wife"
his
married
husband
"wife of"
"married to"
"her husband"
to
her

wife
"his wife"
married
husband
"married to"
"wife of"
"is married"
"is married to"
"her husband"
"(wife"

Lower their influence on Universal Schema

Results

TAC 2016 Component Testing

The recall is low because we don't include Chinese query and Chinese corpus

KB, XLING	Hop0			Hop1			Total
	Prec	Recall	F1	Prec	Recall	F1	F1
USchema	0.475	0.015	0.028	0.071	0.003	0.005	0.021
LSTM	0.421	0.038	**0.069**	0.106	0.018	0.030	**0.057**

SF, ENG	Hop0			Hop1			Total
	Prec	Recall	F1	Prec	Recall	F1	F1
DS	**0.244**	0.168	0.199	**0.500**	**0.013**	**0.025**	0.159
SES+DS	0.229	**0.203**	**0.215**	0.444	0.013	0.025	**0.174**

TAC 2016 Results

LDC Max All Micro Scores

SF	Hop0 (ENG: 350, SPA: 150)			Hop1 (ENG: 185, SPA: 53)			Total
	Prec	Recall	F1	Prec	Recall	F1	F1
2015 system	**0.302**	0.154	0.204	0.081	**0.088**	**0.084**	0.155
2016 system	0.229	**0.203**	**0.215**	**0.444**	0.013	0.025	**0.174**
Spanish	0.1364	0.285	0.187	N/A	N/A	N/A	0.160

2015 KB evaluation							
2015 system	0.227	0.149	0.192	0.038	0.097	**0.054**	0.120

KB	Hop0			Hop1			Total
	Prec	Recall	F1	Prec	Recall	F1	F1
2015 system	**0.390**	0.128	0.192	**0.253**❓	**0.078**	**0.119**	**0.168**❓
2016 system	0.271	**0.154**	**0.196**	0.039	0.071	0.051	0.127
Spanish	0.280	0.121	0.169	0.060	0.048	0.053	0.124

Result Summary

- Compositional Universal Schema (LSTM) improves recall significantly.

- Search engine supervision cleans training data well.

- Universal Schema can be effectively applied to the language with low resources.

- Our 2015 system is scored much better in this year.

  - Need further investigation.

Q & A