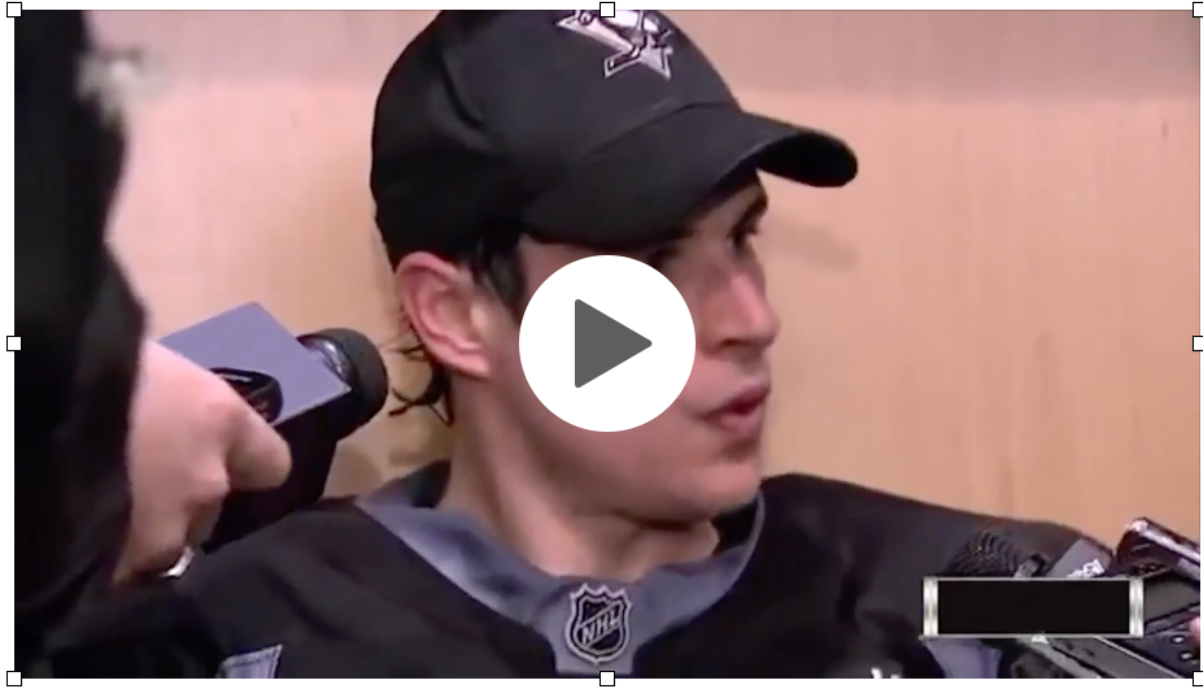

Tie-breaker: using language models to quantify gender bias in sports journalism

—— Liye Fu, Cristian Danescu-Niculescu-Mizil, Lillian Lee ——
Cornell University

#CoverTheAthlete



<http://covertheathlete.com/>

Bias in sports journalism(?)



wimbledon women



In the news



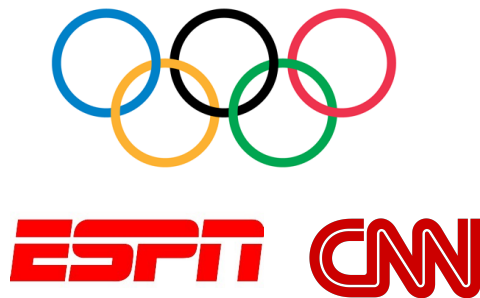
Female Tennis Players Crying 'Fault!' Over Nike's Skimpy Wimbledon Outfit

Inside Edition - 2 days ago

A new dress worn by some **female** tennis stars at **Wimbledon** is getting more attention than ...

Bias in sports journalism(?)

Air-time



[Eastman and Billings, 2000]
[Higgs et al., 2003]

Stereotypical descriptions & Framing



[Angelini and Billings, 2010]
[Kian et al., 2009]

Bias in sports journalism(?)

- *not* in consensus
- subjective manual coding
- targeting only a few specific players?

This work: objectively quantifying gender bias at scale

Data



image source: <http://www.gettyimages.com/event/the-championships-wimbledon-556914085#andy-murray-of-great-britain-attends-a-press-conference-after-losing-picture-id480260312>

Data



Jan, 2007 ~ Oct, 2015

6467 interviews

80000+ Questions

167 female & 191 male players



Data

Q. Can you talk about your next opponent?

A: ...

Q. After practice, can you put tennis a little bit behind you and have dinner, shopping, have a little bit of fun?

A:



<http://www.asapsports.com/>

**... when all is said,
we're done.®**

First attempt: word level analysis

Male players:

clay

challenger(s)

tie

sets

practiced

tiebreaker

maybe

see

impression

serve

still subjective!

height

support

shots

Female players:

yet

new

every

matter

become

meet

winning

type

won

draw

words are ranked in decreasing order of difference in percentage of male and female players who have ever been asked the term.

Game language



The serve-and-volley is being used frequently by Federer and it's enabling him to take control behind his own serve.

Game language

GAME 7: J. Del Potro 4 - 3



J. Del Potro wins the game with a forehand Smash Winner.



40 - 30



J. Del Potro loses the point with a double fault.

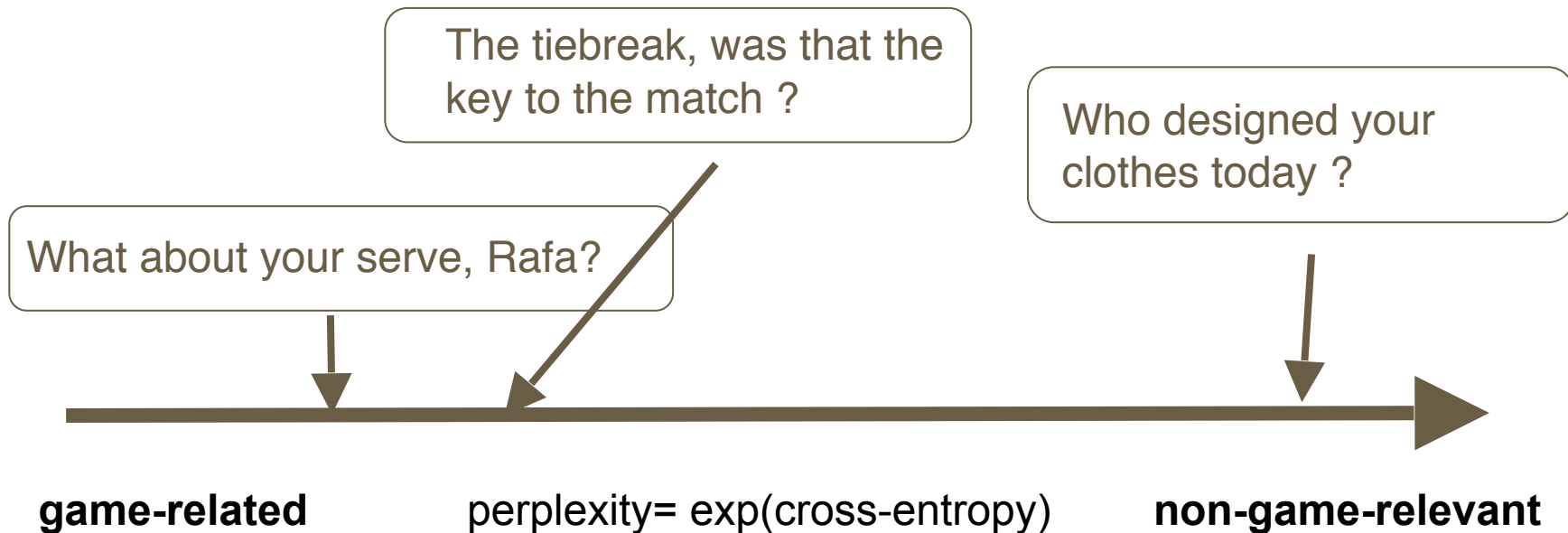
40 - 15

S. Wawrinka loses the point with a backhand Unforced Error.

A gender-balanced set of **3000+** commentaries from

<http://www.sportsmole.co.uk>

Closeness to game language

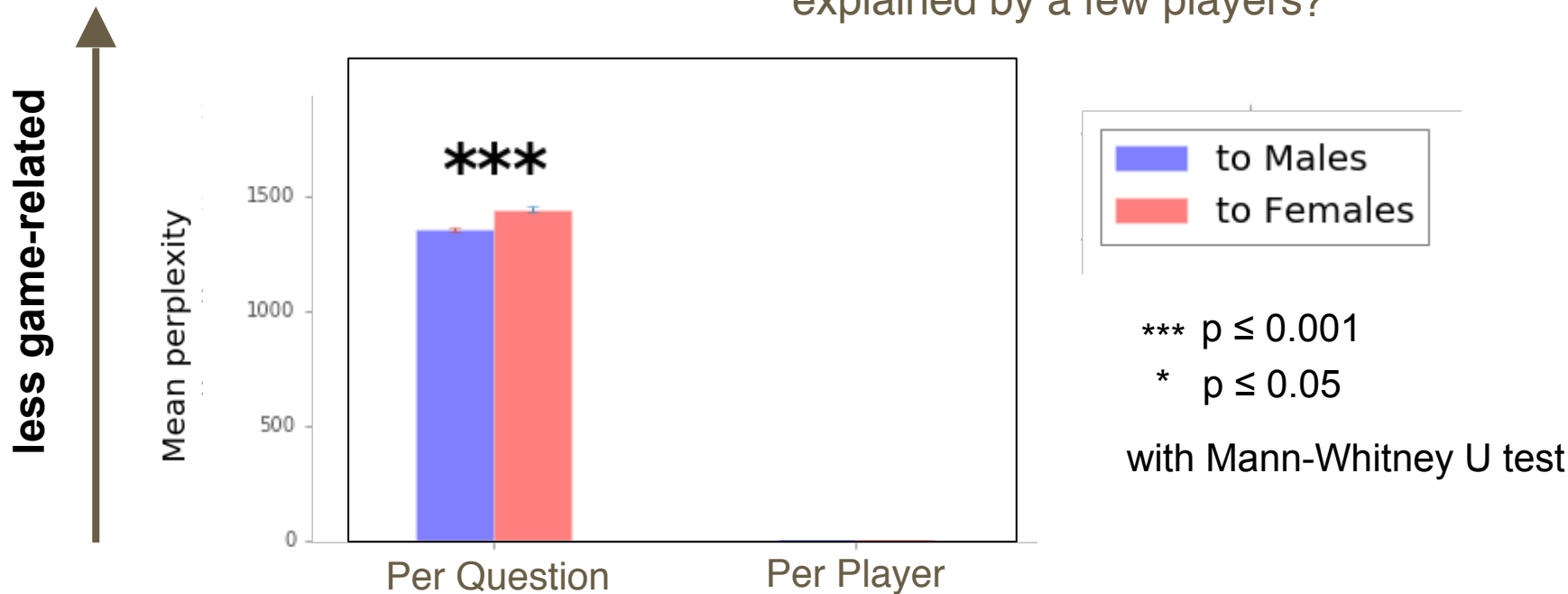


bigram language model trained with KenLM, which uses modified Kneser-Ney smoothing.

Main result

Questions to male athletes are closer to game language

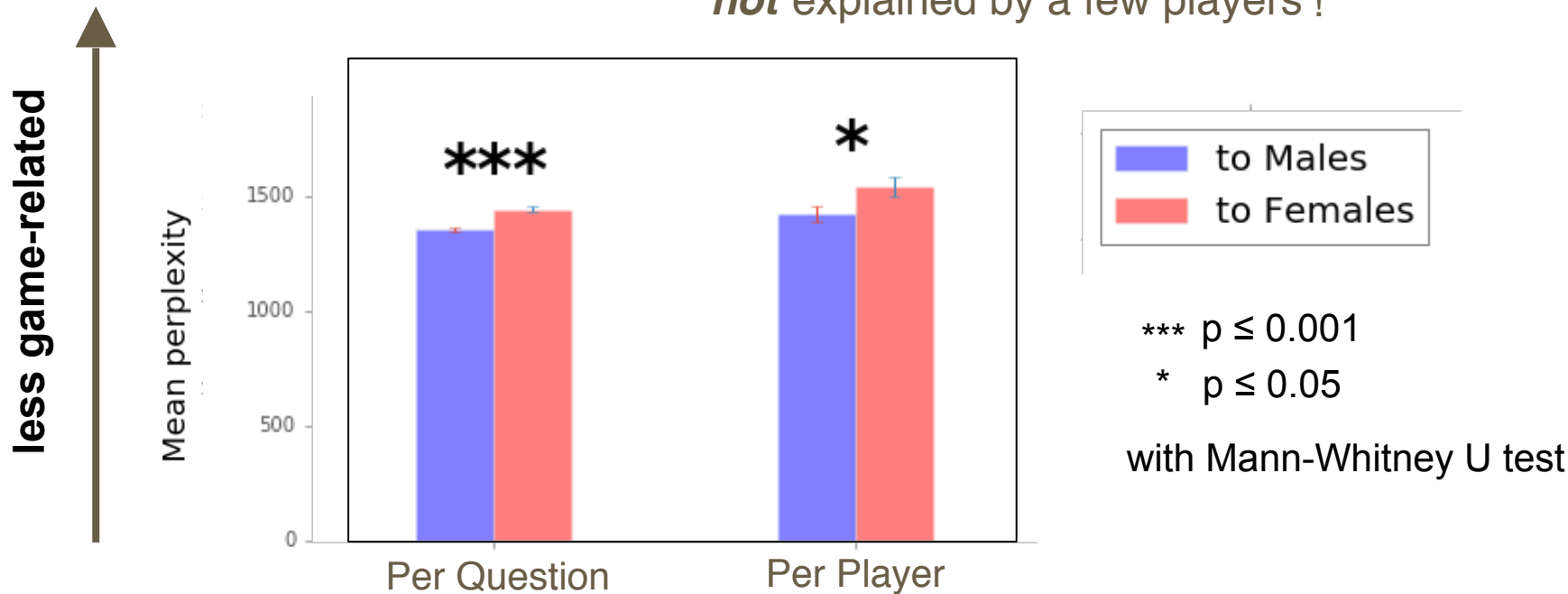
explained by a few players?



Main result

Questions to male athletes are closer to game language

not explained by a few players !




Other experiments



type of question

CENTRE COURT - LADIES' SINGLES				
	Pts	1	2	3
S. Williams (USA) [...]	✓	6	6	
E. Vesnina (RUS)		2	0	

game outcome

RANKING ^	COUNTRY ^	PLAYER ^
1		Novak Djokovic
2		Andy Murray
3		Roger Federer

player ranking

Conclusions

- New language-model-based approach to quantify gender bias
- Questions to male athletes are generally more game-related
- Dataset & our measure can be useful to evaluate other types of bias

Conclusions

- Who asked which question?
- Player responses?
- Other domains?