
Sí o no, ¿què penses?

Catalonian Independence and
Linguistic Identity on Social Media

Ian Stewart
[@alethioguy](https://twitter.com/alethioguy)

Yuval Pinter
[@yuvalpi](https://twitter.com/yuvalpi)

Jacob Eisenstein
[@jacobeisenstein](https://twitter.com/jacobeisenstein)

NAACL 2018



Politics and language

- Language variation often reveals aspects of **social identity**, such as race and gender.
- Local languages have special connection to regional and **political identity**.
- The role of language in political identity expression is magnified during controversy.

How does local language use correlate with political expression on social media?



<http://www.noticiasdegipuzkoa.eus/2018/03/28/politica/intensa-jornada-de-protestas-para-defender-la-republica-catalana>



<https://www.heraldo.es/noticias/nacional/2017/09/22/un-centenar-personas-manifiesta-frente-sede-anc-por-unidad-espana-1198045-305.html>

Study objectives

- Case study: 2017 Catalonia referendum.
 - 92% vote for secession from Spain.
 - Catalan is local language in Catalonia, distinct from Spanish.

RQ1: Is a speaker's **stance on independence** strongly associated with their Catalan usage?

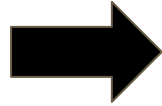
RQ2a: Does Catalan usage vary depending on whether the **discussion topic** is related to the referendum?

RQ2b: Does Catalan usage vary depending on the **imagined audience**?

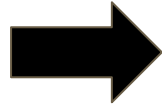
Hashtag collection

**1% Twitter sample
(1/2017 - 10/2017)**

Query

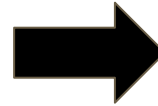


**Referendum seed
hashtags (N=5)**

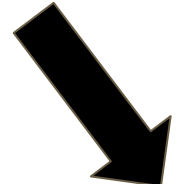
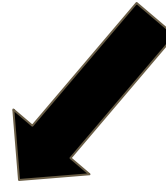


Top 1000 hashtags

Filter



**Referendum
hashtags (N=46)**



**Pro
(N=15)**

**Anti
(N=14)**

**Neutral
(N=17)**

Referendum hashtags

Pro-independence	#1ONoTincPor (18), #1octL6 (184), #CatalonialsNotSpain (10), #CATvotaSí (3), #CataluñaLibre (27), #FreePiolin (293), #Freedom4Catalonia (2), #IndependenciaCataluña (9), #LetCatalansVote (3), #Marxem (102), #RepúblicaCatalana (212), #Spainispain (8), #SpanishDictatorship (9), #SpanishRepression (3), #TotsSomCatalunya (261)
Anti-independence	#CataluñaEsEspaña (69), #DontDUIt (12), #EspanaNoSeRompe (29), #EspañaUnida (4), #OrgullososDeSerEspañoles (55), #PorLaUnidadDeEspaña (2), #ProuPuigdemont (187)
Neutral	#1O (748), #1Oct (1351), #1Oct2017 (171), #1Oct2017votarem (28), #CatalanRef2017 (46), #CatalanReferendum (3244), #CatalanReferendum2017 (72), #JoVoto (54), #Ref1oct (90), #Referéndum (640), #Referendum1deoctubre (146), #ReferendumCAT (457), #ReferendumCatalan (298), #Votarem (954)

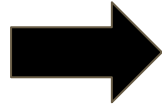
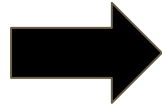
Referendum hashtags

Pro-independence	#1ONoTincPor (18), #1octL6 (184), #CatalonialsNotSpain (10), #CATvotaSí (3), #CataluñaLibre (27), #FreePiolin (293), #Freedom4Catalonia (2), #IndependenciaCataluña (9), #LetCatalansVote (3), #Marxem (102), #RepúblicaCatalana (212), #Spainispain (8), #SpanishDictatorship (9), #SpanishRepression (3), #TotsSomCatalunya (261)
Anti-independence	#CataluñaEsEspaña (69), #DontDUIt (12), #EspanaNoSeRompe (29), #EspañaUnida (4), #OrgullososDeSerEspañoles (55), #PorLaUnidadDeEspaña (2), #ProuPuigdemont (187)
Neutral	#1O (748), #1Oct (1351), #1Oct2017 (171), #1Oct2017votarem (28), #CatalanRef2017 (46), #CatalanReferendum (3244), #CatalanReferendum2017 (72), #JoVoto (54), #Ref1oct (90), #Referéndum (640), #Referendum1deoctubre (146), #ReferendumCAT (457), #ReferendumCatalan (298), #Votarem (954)

Referendum data collection

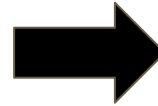
**1% Twitter sample
(1/2017 - 10/2017)**

**Referendum
hashtags (N=46)**



190,060 tweets

**Filter
(high URL)**

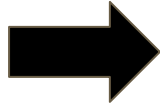


CT

**11,670 tweets
10,498 users**

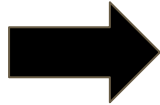
Control data collection

**1% Twitter sample
(1/2017 - 10/2017)**



XT

45,222 tweets



CT users

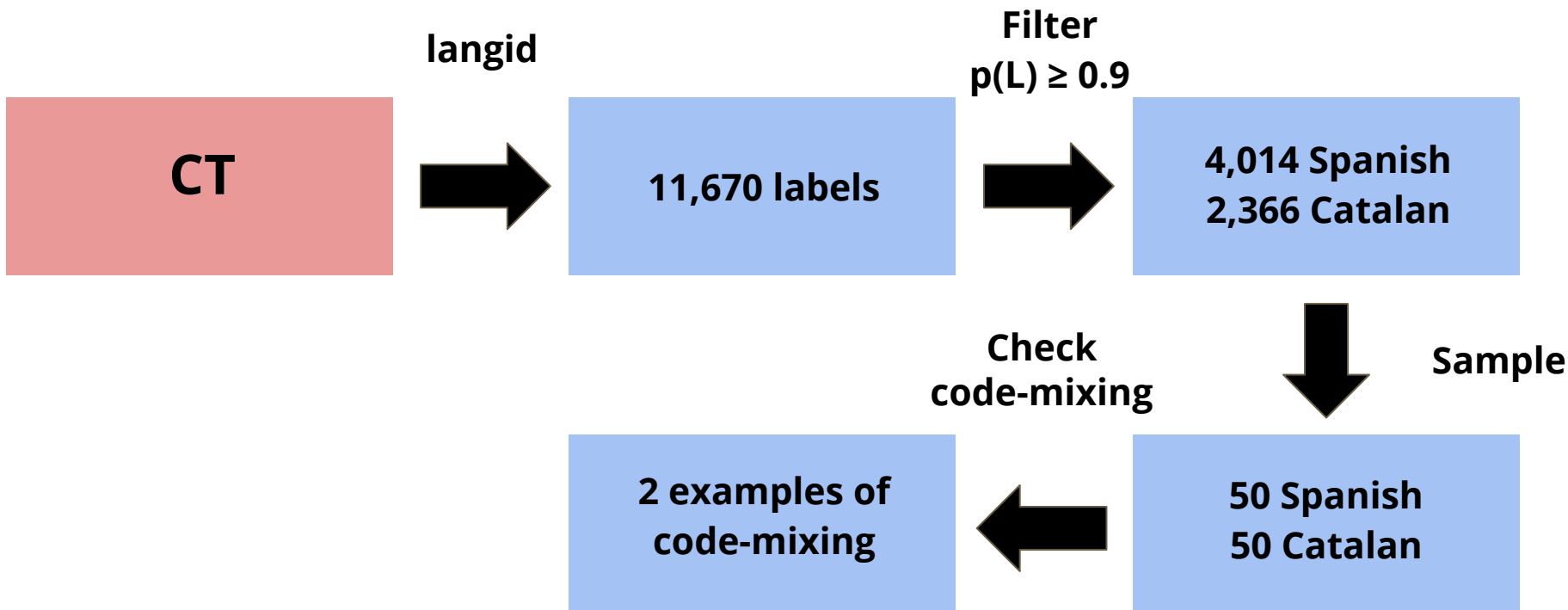
Language identification



Language identification



Language identification



PRO

██████████ · 23 Apr 2017
A punt pel Sí a la [#republicacatalana](#) [#viuSantJordi](#) [#Valls](#) [#esquerra_valls](#)

Ready for Yes to the [#CatalanRepublic](#)
[#ViuSantJordi](#) [#Valleys](#) [#left_valleys](#)

██████████ · 3 Oct 2017
Entiende [@marianorajoy](#) Si tu limitada capacidad de deja. [#CatalunaLibre](#) te manda este mensaje!!

Listen [@marianorajoy](#) If your limited capacity allows it. [#FreeCatalonia](#) sends this message!!

██████████ · 30 Sep 2017
Piolín, no ens oblidem de tu! [#EscolesObertes](#) [#FreePiolin](#) [#NoPassaran](#)

Piolin, let's not forget about you! [#OpenSchools](#)
[#FreePiolin](#)

PRO

23 Apr 2017
A punt pel Sí a la [#republicacatalana](#) [#viuSantJordi](#) [#Valls](#) [#esquerra_valls](#)

Ready for Yes to the [#CatalanRepublic](#)
[#ViuSantJordi](#) [#Valleys](#) [#left_valleys](#)

3 Oct 2017
Entiende [@marianorajoy](#) Si tu limitada capacidad de deja. [#CatalunaLibre](#) te manda este mensaje!!

Listen [@marianorajoy](#) If your limited capacity allows it. [#FreeCatalonia](#) sends this message!!

30 Sep 2017
Piolín, no ens oblidem de tu! [#EscolesObertes](#) [#FreePiolin](#) [#NoPassaran](#)

Piolin, let's not forget about you! [#OpenSchools](#)
[#FreePiolin](#)

ANTI

9 Oct 2017
Replying to [@el_pais](#)
[@AdaColau](#) está **cagadita**, al igual que todos los independentistas.
[#EspanaEnPie](#) [#EspanaUnida](#)

@AdaColau is screwed, just like all the independent-ists. [#SpainOnTrack](#) [#UnitedSpain](#)

4 Oct 2017
[@JoanTarda](#) Ya en tu foto de perfil se ve que eres un descerebrado iluminado. [#CatalunaEsEspana](#) Y si te molesta, vete a Francia.

@JoanTarda Already in your profile picture it looks like you're a brainless visionary. [#CataloniaIsSpain](#) And if it bothers you, go to France.

4 Oct 2017
Me explica alguien eso de la "resistencia pacífica"? [#CatalanReferendum](#)
[#Constitucion](#) [#1Oct](#) [#EspanaNoSeRompe](#)

Will someone explain that about the "peaceful resistance"? [#CatalanReferendum](#) [#Constitution](#) [#1Oct](#)
[#SpainIsNotBreaking](#)

Analysis

- RQ1: Measure difference in Catalan use between:
pro-independence and **anti-independence** Twitter users.
- RQ2: Measure difference in Catalan use between:
(a) **referendum-related** and **non-referendum related** tweets;
(b) **large-audience** and **small-audience** tweets.
- Compare with prior work on political language variation:
(Shoemark et al. 2017).

Comparison



Aye or naw, whit dae ye hink? Scottish independence and linguistic identity on social media (Shoemark et al., 2017)

- Scotland referendum
 - 45% for independence
- Code-mixing (word)
- 30% Scots understanding



- Catalonia referendum
 - 92% for independence
- Code-switching (sentence)
- 80% Catalan understanding

RQ1: analysis metrics

U_{pro} = users with $\geq 75\%$ pro-independence hashtags

U_{anti} = users with $\geq 75\%$ anti-independence hashtags

$p^{(u)}$ = fraction of tweets in Catalan

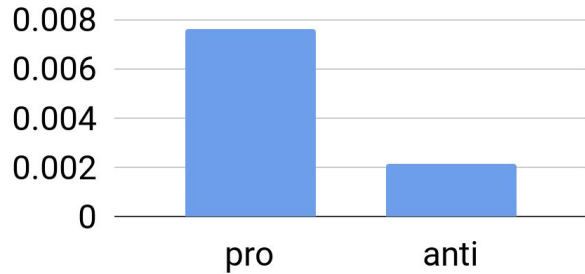
p_{pro} = macro average over U_{pro}

$$d = p_{pro} - p_{anti}$$

RQ1: Do separatists tweet more with local language?

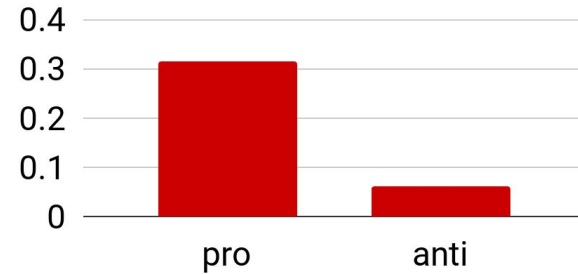
d=0.00555 ($p<0.01$)

Scots usage



d=0.2523 ($p<0.001$)

Catalan usage

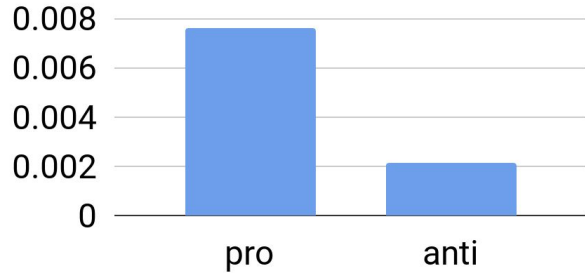


RQ1: Do separatists tweet more with local language?

d=0.00555 ($p<0.01$)

Yes

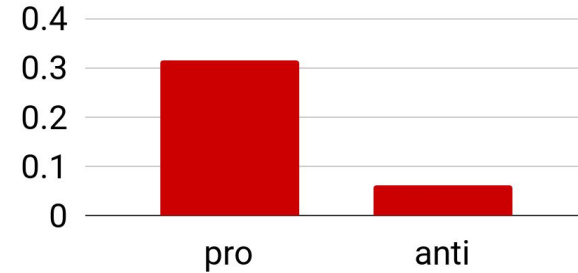
Scots usage



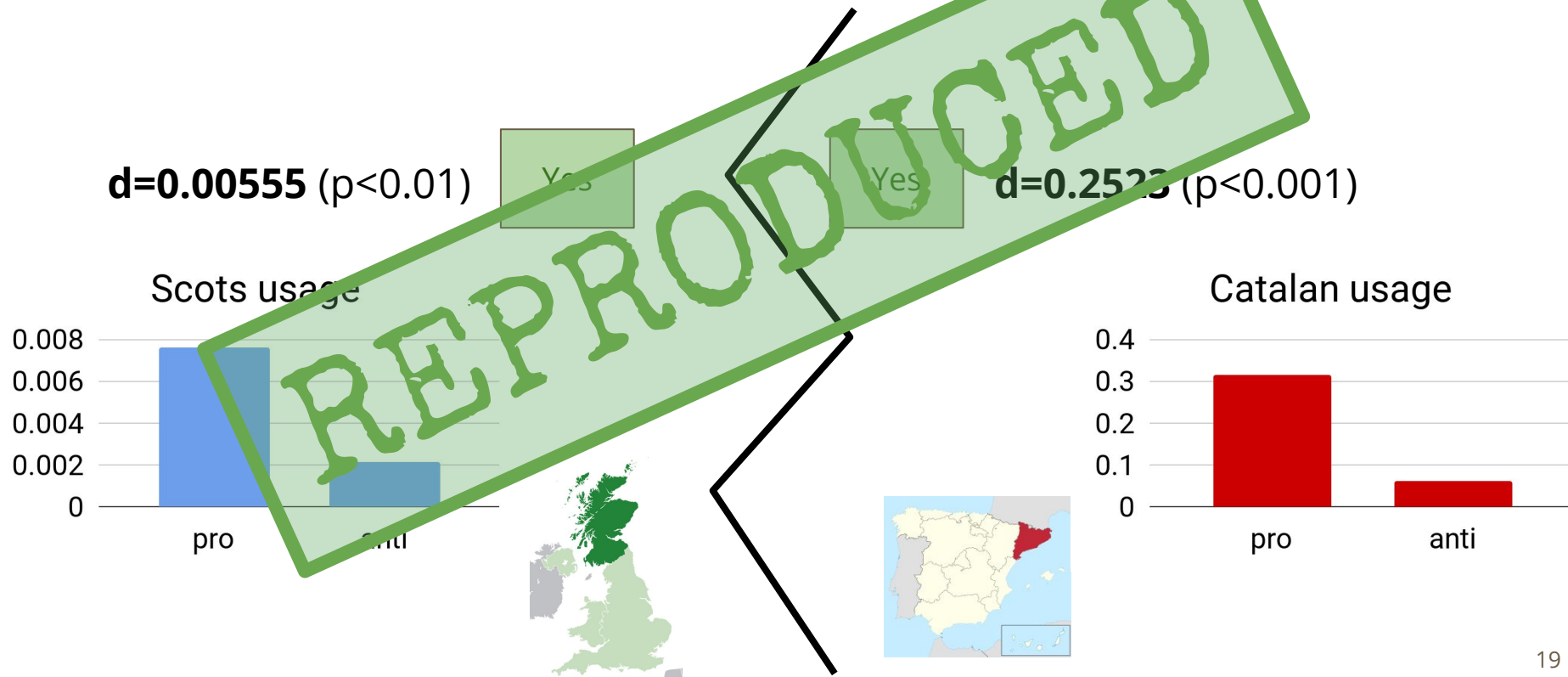
Yes

d=0.2523 ($p<0.001$)

Catalan usage



RQ1: Do separatists tweet more with local language?



RQ2 metrics

$p_c^{(u)}$ = fraction of referendum tweets in Catalan by user u

$p_x^{(u)}$ = fraction of control tweets in Catalan by user u

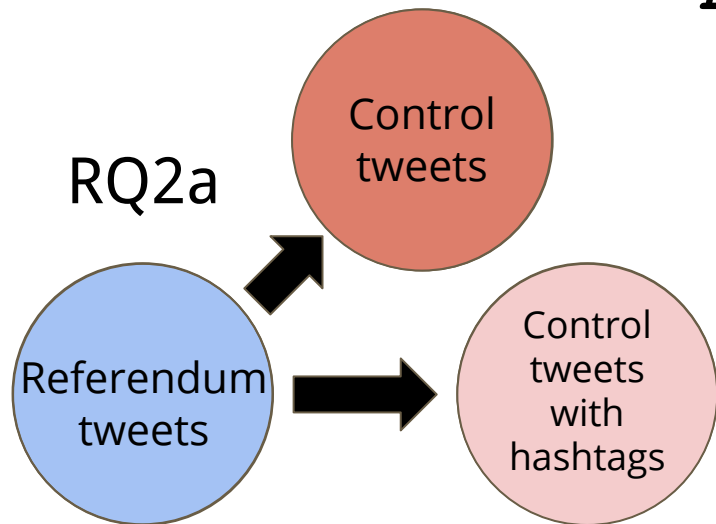
$$d^{(u)} = p_c^{(u)} - p_x^{(u)}$$

RQ2 metrics

$p_C^{(u)}$ = fraction of referendum tweets in Catalan by user u

$p_X^{(u)}$ = fraction of control tweets in Catalan by user u

$$d^{(u)} = p_C^{(u)} - p_X^{(u)}$$

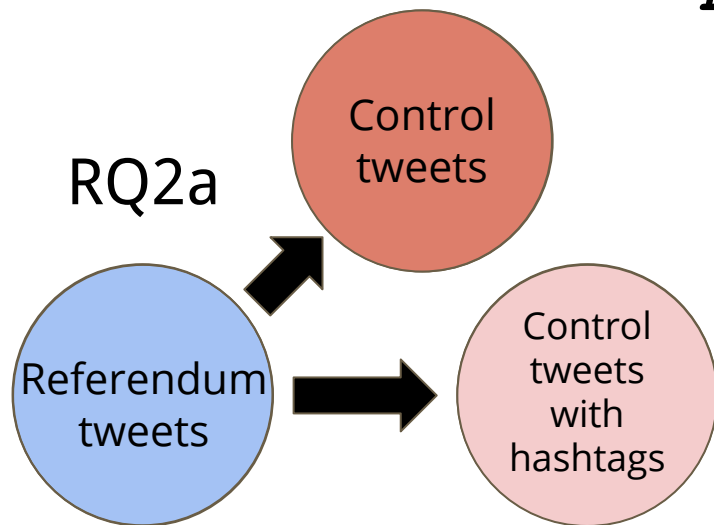


RQ2 metrics

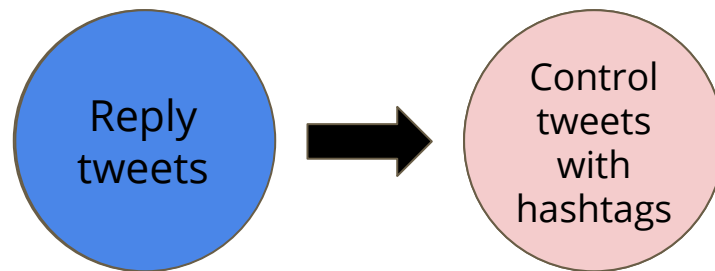
$p_C^{(u)}$ = fraction of referendum tweets in Catalan by user u

$p_X^{(u)}$ = fraction of control tweets in Catalan by user u

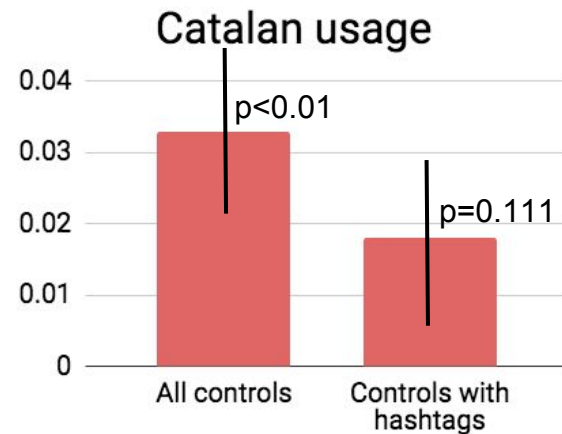
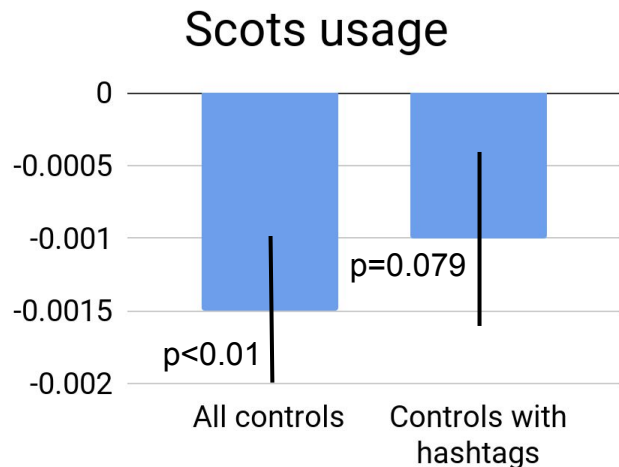
$$d^{(u)} = p_C^{(u)} - p_X^{(u)}$$



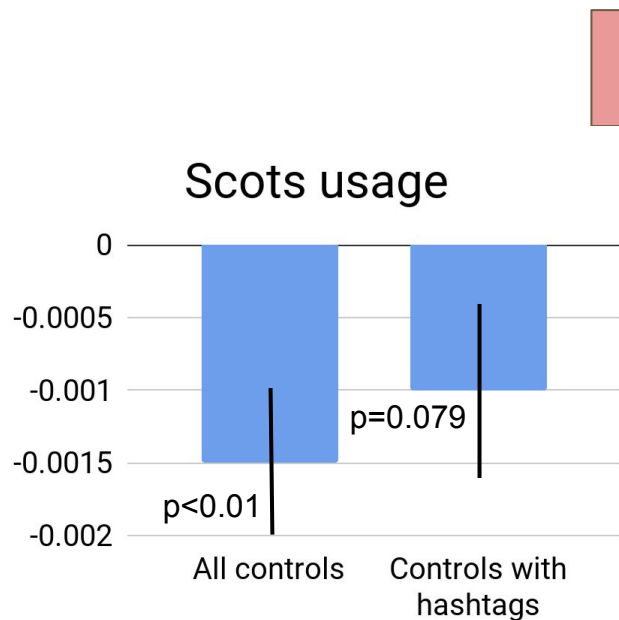
RQ2b



RQ2a: Is local variant more likely in referendum discourse?



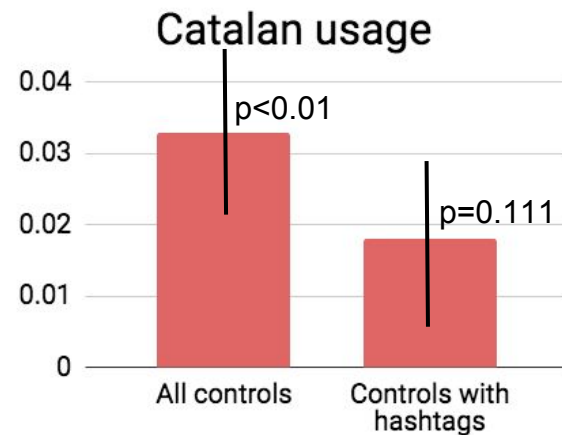
RQ2a: Is local variant more likely in referendum discourse?



No

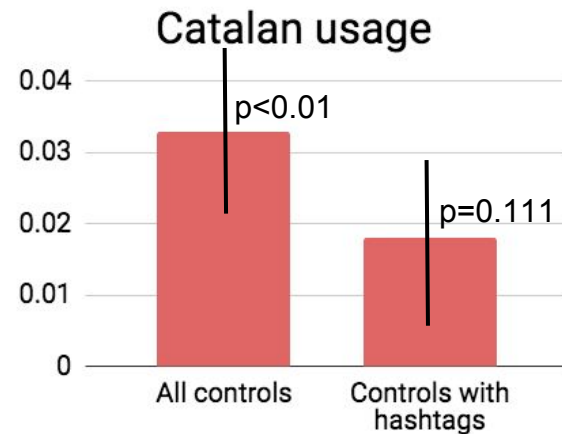
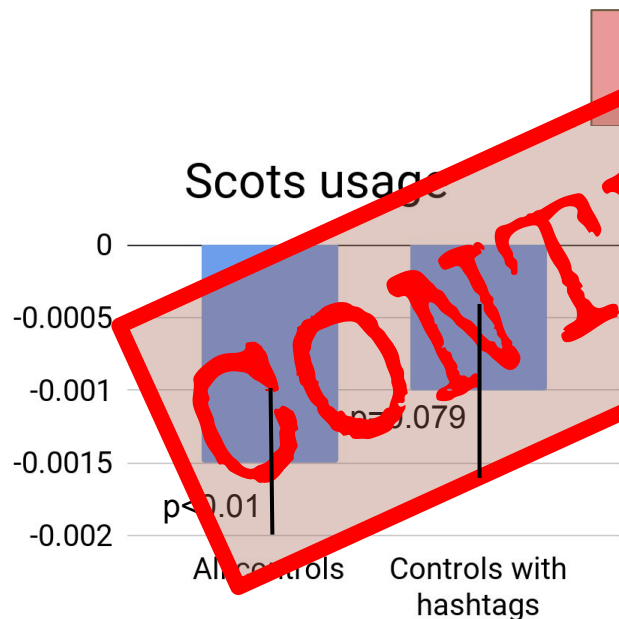


Yes

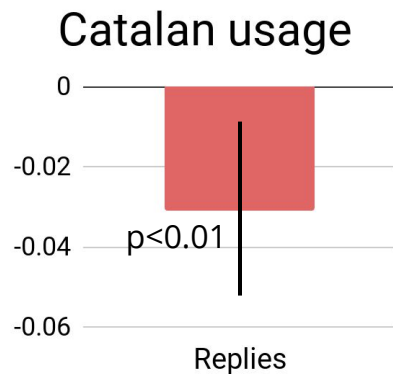


RQ2a: Is local variant more likely in referendum discourse?

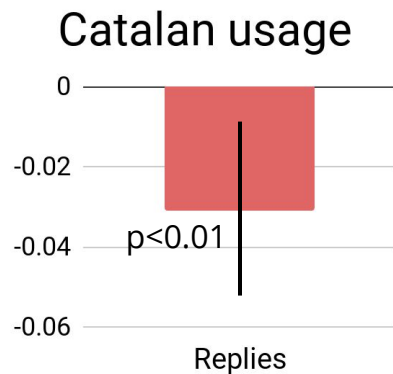
CONTRADICTORY



RQ2b: Is local variant more likely with smaller audience?



RQ2b: Is local variant more likely with smaller audience?



No



@ierrejon What justice are you talking about? What can a JUDGE like this impart?

Conclusions

- Pro-independence tweeters **more likely** to use Catalan than anti-independence tweeters.
- Tweeters **more likely** to use Catalan for broad audience than for narrow audience.
 - Local language less likely to be useful with replies (mostly monolingual Spanish speakers).
- **More extreme** language patterns in Catalonia versus Scotland.
 - Linguistic identity could be more pronounced with code-switching than with code-mixing.
- Future work: political expression in societies with more **evenly distributed** languages (Belgium).

Gràcies! ¡Gracias!

- Questions?

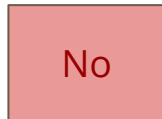
Code: <https://github.com/ianbstewart/catalan>

Extra slides

Experiment 2: Is local variant more likely in referendum discourse?

- All controls

- $d_u = -0.0015$
- $\text{stderr} = 0.0005$
- $t = -2.996$
- $p = 0.0027$



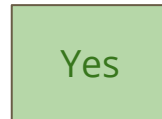
- Controls with hashtags

- $d_u = -0.0010$
- $\text{stderr} = 0.0006$
- $t = -1.758$
- $p = 0.0788$



- All controls

- $d_u = 0.033$
- $\text{stderr} = 0.011$
- $t = 3.02$
- $p = 0.002$



- Controls with hashtags

- $d_u = 0.018$
- $\text{stderr} = 0.011$
- $t = 1.59$
- $p = 0.111$



Experiment 2b: Is local variant more likely with smaller audience?

- Replies
 - $d_u = -0.031$
 - $\text{stderr} = 0.011$
 - $t = -2.79$
 - $p = 0.005$



@ierrejon What justice are you talking about? What can a JUDGE like this impart?

High-level takeaways

- Language choice can be political!
 - Building NLP systems that account for difference intelligently, not necessarily assigning meaning to every local language.
- Use what works: langid.
 - Need to get more creative to handle dialects, actual code-switching.
- Political variation must be contextualized, e.g. with a referendum.
- NLP+CSS researchers should think creatively about other instances of variation beyond typical demographics.
 - Rich opportunities in attitude-based variation, e.g. regional pride.

Seed hashtags

- #CataluñaLibre (PRO)
- #IndependenciaCataluña (PRO)
- #CataluñaEsEspaña (ANTI)
- #EspañaUnida (ANTI)
- #CatalanReferendum (NEUT)