Writing vs. Speech: How do Spanish online videos pronounce gender inclusive language? [*]

Katie Slemp, David Heap, Yarubi Díaz
The University of Western Ontario

1. Introduction

Like most Romance languages, Spanish is traditionally viewed as having grammatically binary gender (Loporcaro 2017). This means that all nouns, whether animate or not, are either masculine or feminine, as are elements that agree with nouns (determiners, adjectives, third-person pronouns). In the case of inanimate subjects, the allocation of gender is arbitrary and adjectives must agree in gender and number when describing any noun (e.g. *las mesas rojas* 'the red tables' F.SG. vs. *el piso negro* 'the black floor' M.SG.). On the other hand, the gender of most animate nouns corresponds to the biological sex of the referent, particularly when denoting humans. (e.g. *una mujer alta* 'a tall woman'.F.SG vs. *un hombre alto* 'a tall man'.M.SG). In recent decades, there has been debate as to how to move towards gender inclusive language, or language that is more representative of all genders. Gender inclusive language, according to Haddad and Baric (2016), is defined as a general intent to favor both men and women in discourse. This definition is binary in nature. For a nonbinary definition, Alpheratz (2018) describes inclusive language as a political act motivated by gender consciousness and the desire to respect symbolic representations and socially excluded categories. Achieving gender inclusive language in Spanish is therefore difficult, because there is a strictly binary system firmly established for many centuries. Furthermore, when referencing a group of both men and women, or when referring to an unknown group, the masculine plural is used almost exclusively (e.g. *los chicos* 'the boys' M.PL or 'the children' M.PL).

In the past few decades, there have been attempts at making Spanish more inclusive, for example by including a doublet when speaking (e.g. *las chicas y los chicos* 'the girls F.PL and the boys M.PL, i.e. kids') and the -@ symbol in writing (*l@s chic@s*, 'the kids' PL). Prescriptivist authorities, such as the Royal Academy of Spanish, consider these innovations to be unnecessary and ungrammatical since the masculine plural form is traditionally considered to denote both the masculine and the feminine. Other graphemes have appeared in writing over recent years that attempt to represent gender neutrality by replacing the normative gender markers -*a* and -*o* with an -*x* marker, and most recently with an -*e*. These graphemes sometimes appear as forms like *todxs* and *todes* ('all' M.F or

'all' N), which are both neutral forms proposed to replace the masculine plural generic *todos* ('all' M.PL).

When the grapheme *-x* began appearing in place of gender-marked *-o* or *-a*, questions arose as to how it would be pronounced, since the sequences [ks] or [ɣs] can only appear in syllable onset or coda position, never as a syllable nucleus in Spanish. This study seeks to explore the written-spoken dichotomy in regard to gender inclusive language in Spanish-language YouTube videos. YouTube videos were analyzed to determine how the creators/speakers orally represented the gender inclusivity that was present in their titles. YouTube permits grammar usage that does not follow established rules, and is also free,open to all users and in the public domain, making it a very useful source for a preliminary study.

2. Methodology

The process of creating a corpus of YouTube videos was lengthy and started with the collection of a judgement sample which was then labelled and sorted to create a corpus of YouTube videos with gender inclusive markers in the titles (-@, -x, -e) and also some forms of inclusivity in the spoken language of the video. We used a variety of search terms and filters which located potentially inclusive terms via the titles of the videos on Youtube. The key words used to compile the sample are shown in Table (1):

(1) Key words used to search for Youtube videos

mexicanes	mexicanxs	mexican@s
Venezolanes	venezolnxs	venezolan@s
colombianes	colombinxs	colombian@s
salvadoreñes	salvadoreñxs	salvadoreñ@s
puertorriqueñes	puertorriqueñxs	puertorriqueñ@s
Peruanes	peruanxs	peruan@s
bolivianes	bolivianxs	bolivian@s
hondureñes	hondureñxs	hondureñ@s
n/a	españolxs	español@s
panameñes	panameñxs	panameñ@s
chilenes	chilenxs	chilen@s
uruguayes	uruguayxs	uruguay@s
paraguayes	paraguaxs	paraguay@s
ecuatorianes	ecuatorianxs	ecuatorian@s
guatemalteques	guatemaltecxs	guatemaltec@s
Cubanes	cubanxs	cuban@s
Dominicanes	dominicanxs	dominican@s
n/a	costarriquencxs	costarriquen@s
n/a	nicaraguensxs	nicaraguen@s
Chiques	chicxs	chic@s
Todes	todxs	tod@s
Latines	latinxs	latin@s
Chicanes	chicanxs	chican@s

N/A indicates forms that are already attested in Spanish and thus not marked as particularly 'inclusive'. Many of the generic forms of nationalities would include -*o* endings, such as *mexicanos*, and are what is often labeled as a "false" generic (since it includes humans of more than one gender). However, a few like *españoles*, are normally attested forms and thus do not overtly express masculine gender due to the -*e* grapheme. About half of these search terms did not yield a result applicable to our corpus. Using these search terms, an initial judgement sample was created corresponding to all the videos with titles containing at least one of the terms in Table 1. Our initial sampling collected as of May 2019 contained 166 videos, but not all of them were useful and further selection criteria (or filters) were needed. The basic judgement criteria were: the video must contain inclusive language in the title, it must contain spoken elements, and the spoken elements must be in Spanish. Some videos (57 in total) had to be eliminated during the sorting step due to trends noticed during the classification process. The types of videos eliminated from the corpus were

either not in Spanish (the -*x* grapheme occurs in Portuguese as well), did not contain spoken elements, were music videos, were instructional videos for inclusive language or were videos intending to motivate the use of inclusive language, publicity videos, music playlists, or were videogamer play. The basic information (title, channel, date of publication, country of origin, and duration) was recorded for the remaining videos (109) and they were then classified as: political, personal vlog, humor, or journalism/news. After the collection, labeling, sorting, and categorization of the corpus, analysis could begin.

3. Results and analysis

The first stage of analysis that took place was categorizing what type of inclusive language was verbalized in the videos of the corpus. We found three possibilities: doublets, epicene -*e*, and the generic feminine. Frequently, these occurred via correction or autocorrection, showing that it is an intentional intervention to restructure the grammatical gender of Spanish. For example, in Argentina when discussing abortion legalization, a student states that it is important for "*nosotras, bueno, nosotres. Hombres y mujeres*" ('we-F. well, we-N. Men and women').

(2) Written-spoken inclusive language Spanish Youtube corpus

	-*x* grapheme	-*@* grapheme	-*e* grapheme
107 videos in corpus	61	34	12
74 videos without spoken inclusive language	40	24	10
33 videos with spoken inclusive language	21	10	2

The order of the doublets was also considered in the analysis, as shown in Table 3:

(3) Spoken inclusive language : doublets (23 occurrences)

	-x grapheme	*-@* grapheme	*-e* grapheme
23 videos with spoken doublets (e.g. *todos y todas* / *todas y todos*)	12	10	1
17 videos with order masculine-feminine (e.g. *todos y todas*)	9	7	1
6 video with order feminine-masculine (e.g. *todas y todos*)	3	3	0

Typically, the spoken doublets occur in the order masculine-feminine, as is shown in the table above. For example, a student body could be referred to collectively in writing as *l@s alumn@s*, which would normally be pronounced in oral discourse with a doublet *los alumnos y las alumnas* ('the students'-M.PL. and 'the students'-F.PL) because the *-@* grapheme does not have an obvious pronunciation, and is intended to indicate both the masculine and the feminine forms. This proposed solution is intended to combat the prescriptively accepted masculine plural form which is traditionally viewed as generic (Real Academia Española 2018), whereby addressing a student body as *los alumnos* ('the students'-M.PL.) can be understood to refer to both males and females. The Real Academia Española (RAE: 2018) states that this is the only valid way to address a mixed gender group, since the masculine plural is understood as encompassing both genders (2018). In recent years, however, the *-@* grapheme in writing and its spoken doublet counterparts have been criticized as not being inclusive of all gender identities, as it still indicates a binary masculine/feminine option (de Onís 2017) and thus excludes reference to those humans who do not identify with one of these two binary or "polar" genders.

(4) Other spoken expression of inclusive language: 10 occurrences

|	*-x* grapheme	*-@* grapheme	*-e* grapheme
9 spoken forms with innovative *-e*	8	0	1
1 generic feminine	1	0	0

The feminine generic (above in Table 4) was used once in our corpus to refer to a mixed gender group. A group of mixed genders is typically referred to using the masculine plural generic, according to prescriptive rules, so using a feminine generic to refer to a mixed group is rare. The spoken forms utilizing the neutral *-e* grapheme have appeared in recent years and are innovative forms that are not part of the current lexicon (e.g. innovative *mexicanes* vs. traditional *españoles*). For example, the word *todes* is said frequently, especially at the beginning of a video, to greet the mixed gender group of possible viewers,

and is more succinct than continuously using doublets. This innovation follows a pattern of existing epicenes words, like *estudiantes* ('students'-PL.), which are only seen as gendered when the article or modifiers are used (e.g. *los estudiantes nuevos* 'the new students' M.PL. and *las estudiantes nuevas* 'the new students' F.PL.).

3.1. Results by country, year, and theme

The type of oralized inclusive language in different videos was classified according to the country of origin of the speaker, see Table (5):

(5) Types of inclusive language according to year of publication
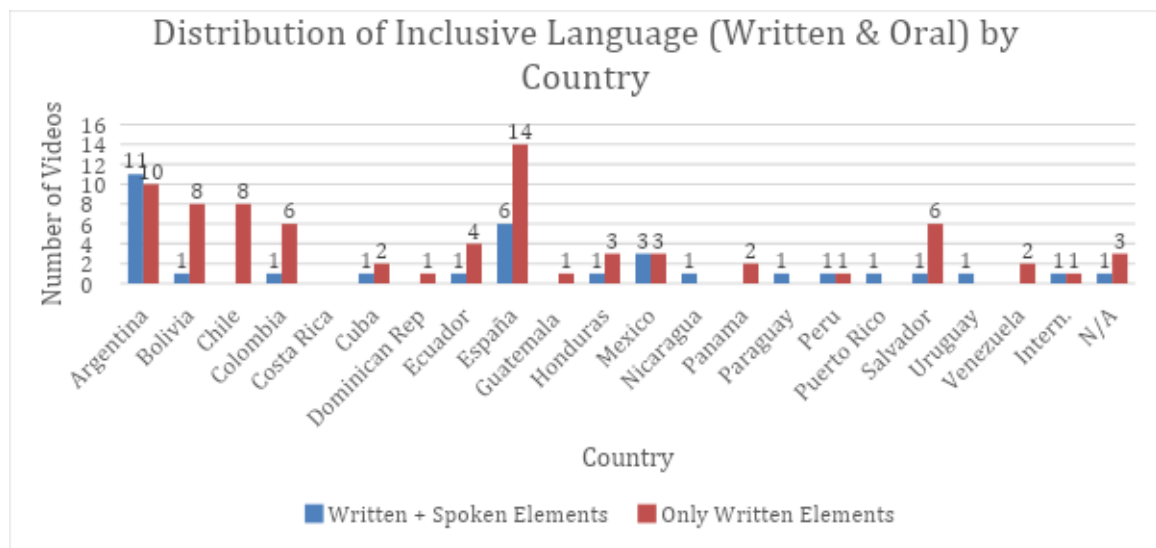
Above, N/A indicates that the country of origin is not stated explicitly or implicitly in the video. Spain and Argentina are both hot-spots for written and spoken inclusive language. The difference between the two countries in our corpus is that Spain inclusive language users prefer to express gender inclusive language only in writing more often than in writing and speech, while inclusive language users in Argentina incorporate written and spoken elements almost equally, but vocalize inclusive language more than twice as often as in Spain.

Additionally, the type of oral and written inclusive language was recorded according to the year the video was published to examine the time trend, as shown in Figure (6):

(6) Spoken elements versus the written elements from Tables (2), (3), and (4) above:
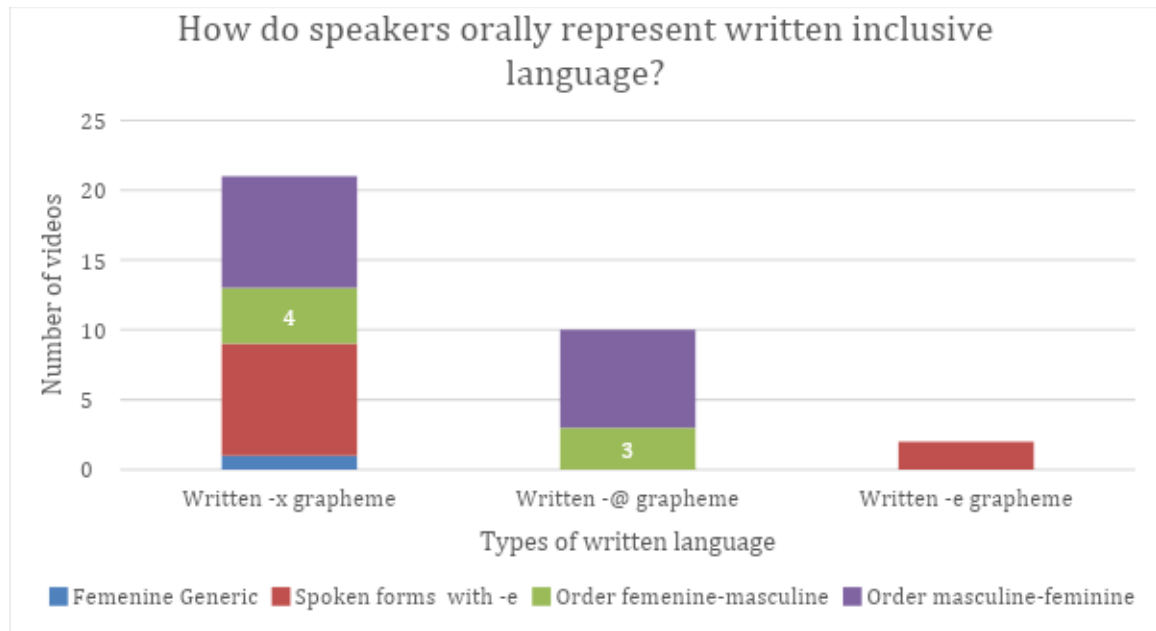

There is more than one way to orally represent written inclusive language. The -@ symbol is represented in speech only by oral doublets, largely in the order of masculine-feminine, while written forms with the innovative epicene -*e* only correspond to oral forms including the same innovative epicene marker. The -*x* grapheme is much more varied in its oral expression. There is approximately the same breakdown between doublet order for -*x* forms as for the -@ symbol, but there is also almost the same number of innovative -*e* forms. There is also one occurrence of feminine generic use. It is not entirely unsurprising that there is such a variety for the -*x* grapheme's oral expression, since /ks/ cannot appear as a syllable nucleus in Spanish and is therefore more varied. In fact, not one video contained a form with /ks/ in syllable nucleus position, regardless of what written inclusive element appeared in the title. The -*x* as an option also began later than the -@, which also may contribute to the variation seen.

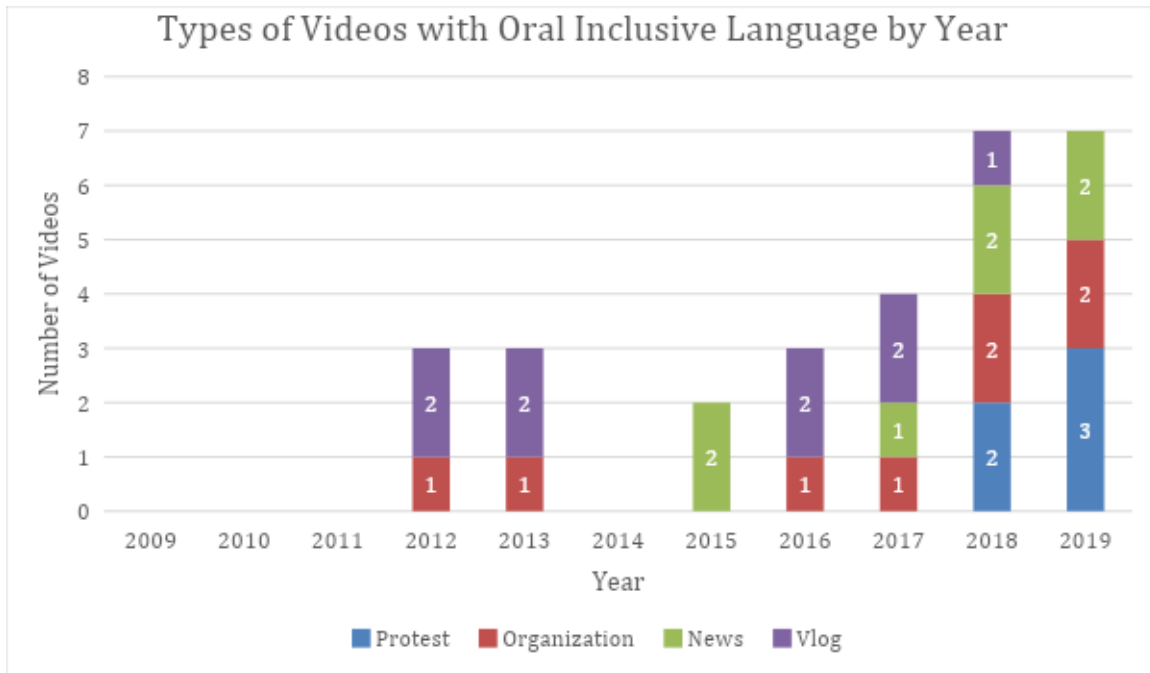(7) Written inclusive language according to year
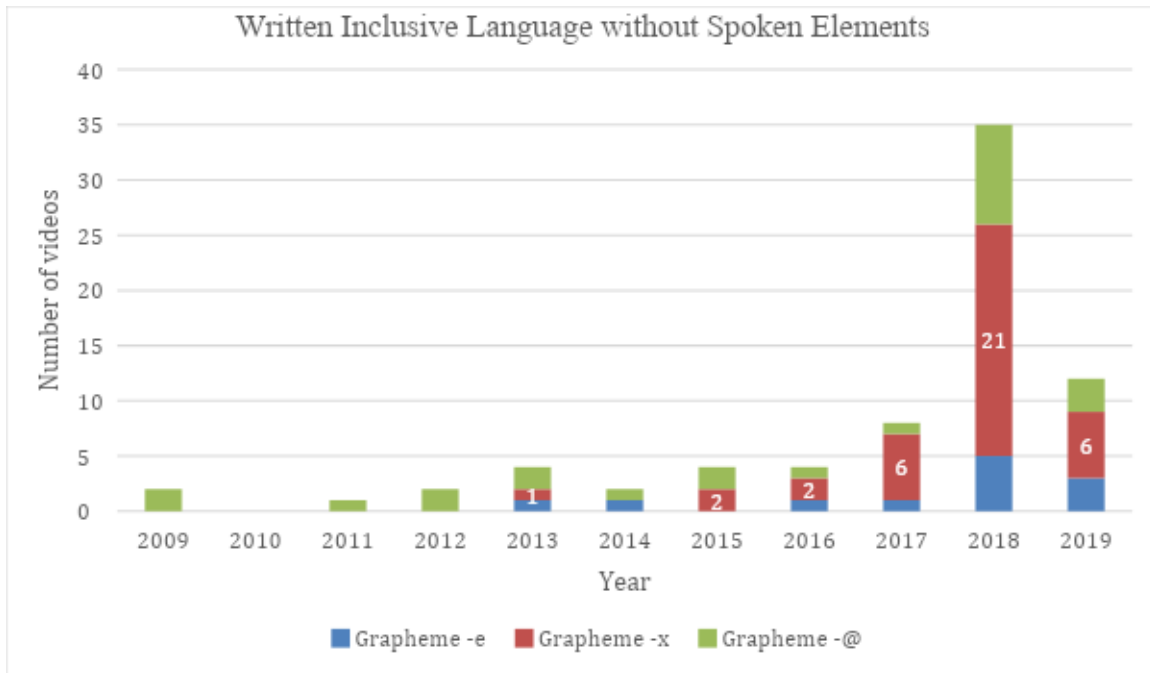
In general, the appearance of inclusive language in the titles of YouTube videos has increased over time, with a large increase between 2017 and 2018. The corpus included videos published from January to May, 2019, which may explain why there are fewer videos in 2019 compared to 2018, but the first five months of 2019 still saw more videos with gender inclusive written elements than the whole of 2017. There is no real explanation for the lack of videos in 2010, but numbers in the first few years barely exceed zero in any case, so initial trends are sporadic. The -@ symbol has been used for a decade, with the earliest occurrences appearing in 2009. The -*x* and -*e* graphemes both appear in the year 2013, but do not consistently outnumber the use of -@ until 2017. The -*x* grapheme appears more often than the -*e* grapheme in all years where they are both present.
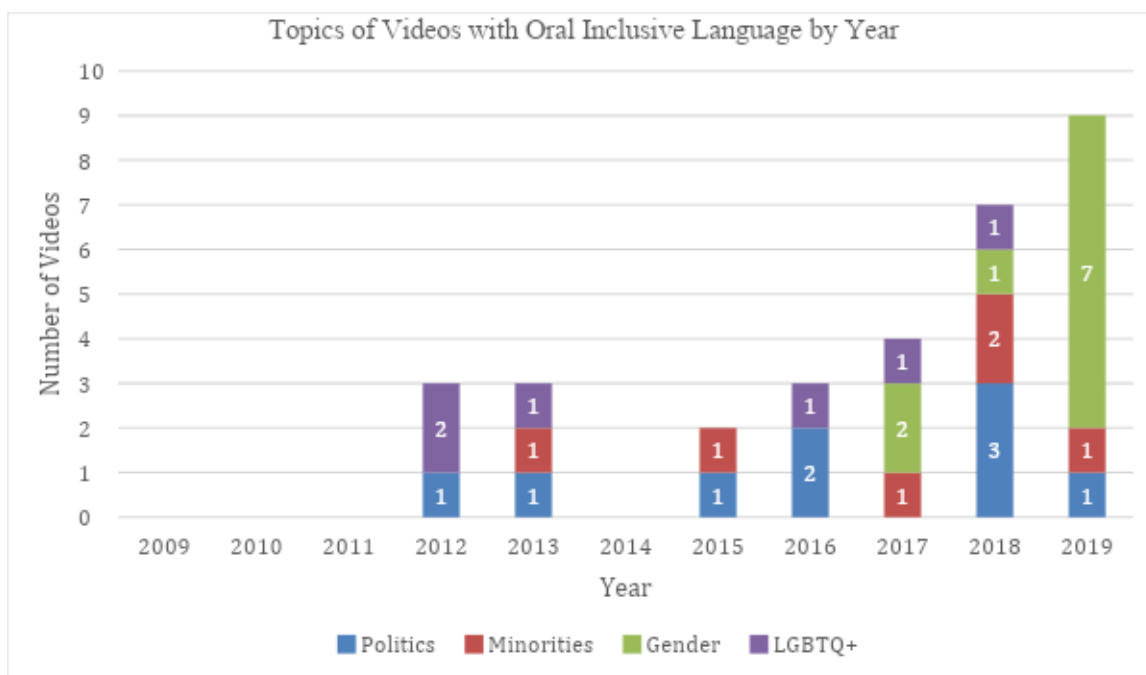
(8) Oral inclusive language according to year
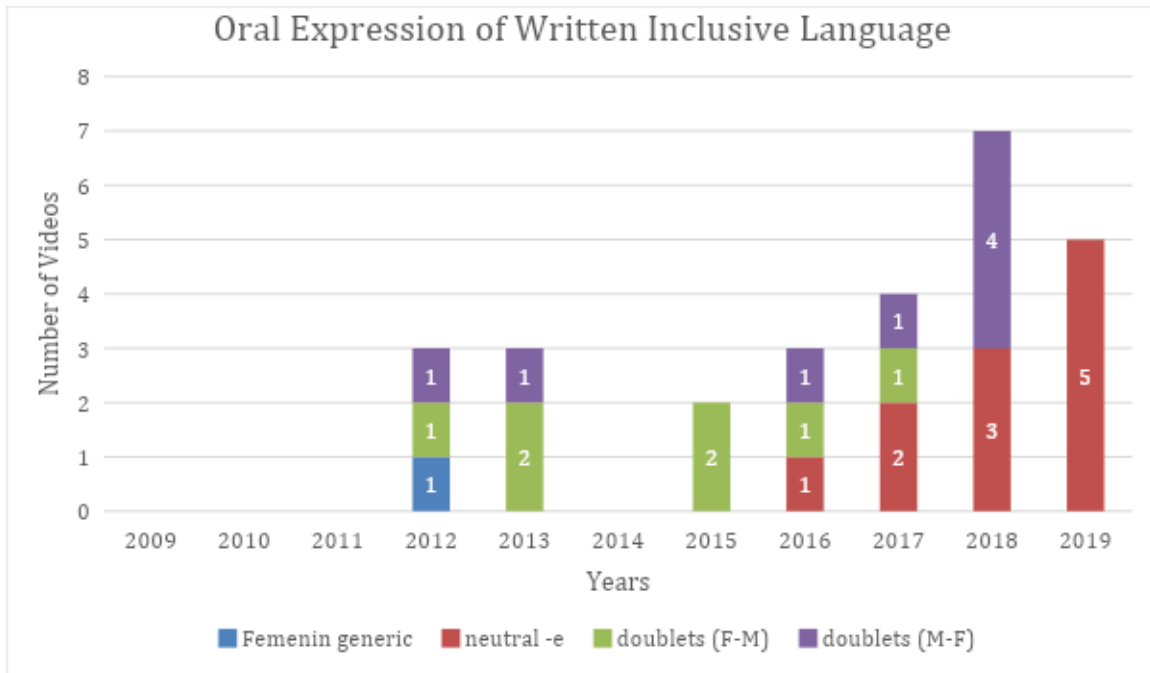
Oral expression of inclusive language starts later than written expression in the corpus compiled, in 2012 (rather than in 2009 when written gender inclusive elements begin to appear). The sole occurrence of a feminine generic occurs in 2012. The remaining inclusive speech is doublets, until the utterance of the neutral -e in 2016. The use of the neutral -e increases each year, but does not outnumber the use of doublets, although it does appear equally in 2017. The use of spoken gender inclusivity increases each year, unlike the sharp increase in 2018 and subsequent decrease in 2019 shown for written elements above in Figure 7. This means that although there was a decrease in written inclusive elements from 2018 to 2019, oral inclusive language has increased, particularly with the incorporation of innovative -e forms, which in both 2018 and 2019 were almost equal to the use of spoken doublets.

(9) Type of video according to year

Types of Videos with Oral Inclusive Language by Year

We can also classify gender inclusive language according to the type of video in which we found it (Figure 8). The types of videos containing spoken inclusive language have also diversified over time. Originally, inclusive language appeared in organizational videos and personal vlogs. In 2015, inclusive language begins to appear in the news and later in 2018 appears in protests. Over time, the types have diversified, but the inclusive language still appears in vlogs and organizational videos where it began.

(10)   Topic of video according to year

We can also classify the corpus according to the themes or subjects people were approaching when adding inclusive language in their speech. There are a select number of topics that appear in videos containing inclusive language. The first topic to appear is unsurprisingly LGBTQ+ in 2012. Gender is the last topic to appear in 2017 but by 2019 becomes the dominant topic. It is surprising that the topic of gender appears much later in time than the other topics as the very subject of inclusive language pertains precisely to the subject.

4. Discussion and conclusion

While our sample is small, it suggests that gender inclusive language is increasing year by year in Spanish, both in writing and in speech. In a YouTube corpus, speakers can freely represent themselves and their identities, without the worry of speaking or writing "correctly". In writing, inclusivity has expanded markedly in the past few years. Inclusive speech is also on the rise, but not on the same scale. It seems that there is not yet a consensus on how to represent gender inclusivity both orally and in writing. The diversity in the attested types of gender inclusivity represent the movement itself: an intervention from the "bottom up" without any unified institutional authority. It represents a desire on the part of some people that the language should change, but with different answers on how to do it best. Despite the attested variation, the innovative gender neutral forms ending in -*e* seem to be the most promising way to include all gender identities and escape the binary system traditionally utilized in Spanish.

The use of these inclusive language markers in our corpus leads us to offer some observations about the graphemes -*x*, -*@*, and -*e*. We noticed that the use of the grapheme -*x* continues to be transgressive in normative Spanish, a conflict that we observe in the different oral manifestations that are associated with this grapheme in our corpus. We believe that two positions converge in its use: (1) the rejection by people who use inclusive language, considering that "the signifier symbolizes linguistic imperialism, poses pronunciation problems, and alienates non-English-speaking immigrants," (from Onis, 2017: 79) and (2) as Roy Perez states, reaffirms the use of the -*x* for its transgressive sexual, gender, and language politics. Perez states, "If we understand language as one medium among many for making political interventions, I think the instability of the -*x* is very useful (more than proper Spanish syntax)." (Roy Pérez, in de Onis, 2017: 82). In other words, the complexity of the use of -*x* in Spanish speech is confronted by the symbolism of transgression of the traditional form.

Concerning the grapheme -*@*, it does not exist in the Spanish alphabet, so without any written or oral assigned norm, we observe that the users did not try to produce it orally with any vocalic sound even when it is placed in a vocalic position. There was also no combination of vowels associated with this symbol, neither /-ao/ nor /-oa/, etc. The speakers of these videos generally used doublets to oralize the words with -*@*. We also noticed that the –*@* was the inclusive language marker least used in the videos, and that its use is decreasing over time. The -*@* symbol is commonly used to denote both the masculine and feminine (i.e. binary) genders, whereas the -*x* came about to show more flexibility than was possible within the strictly binary system.

Regarding the marker -*e*, we can see that this form has several advantages. At the morphological level, -*e* is already an existing desinence in the nominal system of Spanish (e.g., *estudiante 'student', grande 'big'.* M.F. SG). Likewise, it can be used in the plural (e.g. *estudiantes 'students'*, grandes 'big'.M.F.PL.) while respecting Spanish morphology and phonology. On the phonetic level, it seems to be also more advantageous than -*x* pronounced /-ɛks/, since Spanish prefers open syllables, and consonant clusters never appear in the final syllable position. These arguments favour the use of -*e* over -*x* (given that /- ks #/ is an impossible coda in Spanish) and over -*@*, which has no obvious pronunciation in Spanish: both factors likely contribute to the increasing use of this form has in recent years.

Despite the increase in the use of inclusive forms, the results indicate that people continue to use the masculine generic: 78/108 of the initial sample of videos containing written inclusive written elements in their titles did not present any oral mark of inclusive language. It is impossible in most cases to know if the singular or plural masculine forms used on our corpus were referring to only men or if the users were intending to include binary genders or even non binary genders with this form. In the 30 videos where oral marks of inclusive language were observed, "generic" masculine forms are always present as well.

From the traditional point of view, the Royal Spanish Academy (Real Academia Española or *RAE*) had been very firm about its posture against inclusive language, since "the doctrine of Spanish" has been constituted from an encoding "formed over the centuries". Some official tweets of the institution can show the actual position they defend:

"So-called inclusive language involves artificially altering the functioning of the morphology of the genre in Spanish under the subjective premise that the use of the generic masculine makes women invisible" (December, 7, 2018, RAE tweet, our translation). Answering questions from users, the Academy remarks: "It is not permissible to use the letter *-x* or the *-e* as a mark of gender. It is also unnecessary, since the grammatical masculine works in our language, as in others, as an inclusive term to refer to mixed groups, or in generic or unspecified contexts." (Twitter, July, 19, 2018, our translation) The RAE criticizes not only the use of these forms, but also the people proposing these changes, arguing that they have no right to propose any changes to the Spanish language: "It is absurd and dictatorial, that different groups -feminists, regional or ethnic - pretend, or even demand, that the RAE incorporate this or that word of their liking, delete from the dictionary that other one of their displeasure, or 'consecrate' the use of any nonsense or mockery that these groups consider pleasing." (Marías, 2018, our translation)

This rejection by academic institutions can be associated with important controversies. In fact, how to handle inclusive language is quickly becoming an important issue, having consequences even in workplaces. For example, a teacher in Mendoza (Argentina) was recently transferred out of her position as vice-principal for using inclusive language in a public setting, thus offending parents (Página 12, 2019). The use of inclusive language in classrooms has been generating this effect in other institutions where some professors report being expelled for such practice. (Perfil, 2018). However, not all academies show the same posture regarding these innovations. For example, the Argentine Academy of Letters states that there is no need to battle or condemn what is a rhetorical phenomenon, not a grammatical one:

> These different proposals (the doublets, the *-@*, the *-x*, the *-e*) are resources for public discourse which have the purpose of denouncing and highlighting injustices in society. That is to say, they are not grammatical phenomena, but rather rhetorical (and very powerful ones), since they are used in order to create an effect, in those who read or listen, to raise awareness about a social and cultural problem. Therefore, condemning its use with grammatical arguments would be the equivalent of condemning any of the metaphors used daily in politics, simply because, in this case, it is a non-traditional operation on gender and its morphology, without too much background in history. At the same time, demanding its use would be the same as forcing someone to adopt a particular political idea, a practice that takes over its power, since the most transformative ideas are always those that are voluntarily adopted, not those that are imposed (Academia Argentina de Letras, 2019, our translation)

In the specific context of language diachrony, our data suggest the possibility of a phenomenon of change where the written form precedes the oral use. For this reason, inclusive language represents an interesting context to study emerging written-to-oral changes. In a more general sense, our approach does not attempt to fix the relevance of the

study of inclusive language based solely on the increase or decrease of its use, nor on the fact that it meets criteria to be classified as linguistic change. We consider, as Kalinowski states, that "Inclusive language does not measure its success by the degree of grammaticalization it acquires, that is, if it achieves a linguistic change, but in achievements around very serious problems that women suffer in society." (Kalinowski 2019, our translation). In this way, as these kinds of social problems exist, we expect speakers will continue proposing language forms that oppose those injustices, expressing the need for equality in the ways we found in our corpus, or even in new ways that may also defy traditional normative usage in Spanish.

References

Academia Argentina de Letras, 2019. La lengua en el centro de un debate social: el caso del lenguaje inclusivo. http://www.aal.edu.ar/?q=node/606

Alpheratz. 2018. Français inclusif : Conceptualisation et analyse linguistique. Dans *SHS Web of Conferences.* (46). 1-21.

Bosque, Ignacio. 2012. Sexismo lingüístico y visibilidad de la mujer: Informe de la Real Academia Española. http://www.rae.es/sites/default/files/Bosque_sexismo_linguistico.pdf

Flaherty, Mary. 2001. How a Language Gender System Creeps into Perception. *Journal of Cross-Cultural Psychology* 32(1), 18–31.

Haddad, Raphaël et Carline Baric. 2016. Manuel d'écriture inclusive. Faites progresser l'égalité des femmes.hommes par votre manière d'écrire. Directed by Raphaël Haddad, Founder and Associate Director. Mots-Clés. http://www.ecriture-inclusive.fr

Kalinowski, Santiago. 2019. Lenguaje inclusivo: cambio linguístico o cambio social. En *Revista del Colegio de Traductores Públicos de la Ciudad de Buenos Aires.* N.º 141. April-June 2019, 53-55 (https://issuu.com/revistatrad/docs/revista141).

Marías, Javier. 2018. "No esperen por las mujeres". *El Pais*. May 11, 2018 https://elpais.com/diario/2008/05/11/eps/1210487217_850215.html

de Onís, Catalina (Kathleen) M. 2017. "What's in an "x"?: An Exchange about the Politics of "Latinx"." *Chiricù Journal: Latina/o Literature, Art, and Culture*, vol. 1, no. 2, 78-91.

Página 12. June 26, 2019. Una vicedirectora sancionada por saludar a "todos, todas, todes" en Mendoza https://www.pagina12.com.ar/202492-una-vicedirectora-sancionada-por-saludar-a-todos-todas-todes.

Perfil. 2018. October 30, 2018. El profesor de Historia que enseña con Lenguaje Inclusivo y habla de "les romanes". https://www.perfil.com/noticias/sociedad/el-profesor-de-historia-que-ensena-con-lenguaje-inclusivo-y-habla-de-les-romanes.phtml

Real Academia Española. 2018. Libro de estilo de la lengua española según la norma panhispánica. Madrid: Espasa