
“The Effect of Competition on Language Diversity in the Movie-Theatre Industry”

Bernat Mallén Alberdi

UBIREA

**Institut de Recerca en Economia
Aplicada Regional i Pública**
UNIVERSITAT DE BARCELONA

WEBSITE: www.ub-irea.com • CONTACT: irea@ub.edu

**AQR**

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • CONTACT: aqr@ub.edu

Universitat de Barcelona

Av. Diagonal, 690 • 08034 Barcelona

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

IREA Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. For that reason, IREA Working Papers may not be reproduced or distributed without the written consent of the author. A revised version may be available directly from the author.

Any opinions expressed here are those of the author(s) and not those of IREA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

Abstract

In this paper I investigate the effect of competition on language diversity in a cultural market, the movies market, in which language is a relevant characteristic of the good. I analyse the case of the bilingual region of Catalonia to empirically test the effect of competition in two stages of the supply chain – the distribution and the exhibition – on the availability of films in the weaker language. I create a unique data set of all the screenings in the region over 10 months from different sources using advanced web-scraping techniques. I find that the concentration at the distribution level reduces the percentage of films in Catalan by 4.04 percentage points compared with the counterfactual of perfect competition. The effect of the concentration at the exhibition level is not significant. This implies that without such market failure, the total supply of films in Catalan would be 96% greater. I also look for heterogeneous effects disentangling two types of audiences: children-targeted films and adult-targeted films. I find that children have higher preference intensity over the language because the market is more responsive to them; the concentration at the exhibition level matters when it comes to this type of consumer.

JEL Classification: D43, L13, L82, Z13.

Keywords: Language diversity, Movie theatres, Dubbing, Bilingualism, Cultural market.

Bernat Mallén Alberdi: AQR-IREA, University of Barcelona. ORCID number: 0000-0003-3779-429X. Email: bernatmallen@ub.edu

1. Introduction

Cultural goods (such as books, films or theatre) and media products (newspapers, TV and radio) deliver their content to the consumers through a particular language. The existence of multilingual local markets makes the choice of language a difficult and sensitive decision for firms (Caminal, 2010). In fact, it has been estimated that more than one-half of the world's population speaks more than one language (Tucker, 2001). The language of these cultural goods will be a determinant factor for the decision of the consumer. For monolingual consumers, the decision will be straightforward: they will consume the cultural good which is in their language. For bilingual consumers, the decision will be more complex.

Economists have changed their view on language. Classically, it was viewed as a mere tool of communication. In this case, the bilingual consumer's decision would depend on the price and the quality of the cultural good; for identical cultural goods with different language options (e.g. a film offered in different versions), the bilingual consumer would choose the one with the lowest price. Nonetheless, nowadays language is no longer regarded by economists as a neutral tool of communication (Caminal & Di Paolo, 2019; Ginsburg & Weber, 2011). The efforts to promote minority languages, the resilience of linguistic groups, the multilingualism in supranational organizations (e.g. EU Parliament) are all examples of the existence and relevance of the subjective dimension of language, which goes beyond the communicative benefits. Such dimension shall not be neglected and actually plays a role in the decision of the bilingual consumer, who has to balance the preference for their own language¹ with the other characteristics of the cultural goods. Some kinds of linguistic preferences have already been introduced in a variety of economic frameworks. See, for example, Grin (1992), Wickström (2005), Mèlitz (2012) and Caminal (2010); the latter is the closest to this paper from a theoretical perspective. It extends the spoke framework (Chen & Riordan, 2007) by adding an additional dimension of product differentiation: language. In Caminal's model, consumers may trade off a good match in terms of content against a good linguistic match for these cultural goods. This paper aims to empirically explore an actual cultural market, the cinema, in a bilingual local market. To the best of my knowledge, it is the first attempt to empirically analyse a cultural market focusing on the linguistic preference and its intensity among consumers.

The movies market presents the perfect set-up to explore the choice of the product's language by firms. As we can see in many countries, offering another language version rather than the original (that is, dubbing) is a common practice, especially when the knowledge of the original language is not generalized among all potential local consumers. Distributors have the rights of film distribution in the country, and they can decide to offer different versions. If one version does not exist, they need to incur the costs of dubbing so as to have this version available. In a second stage, local exhibitors, who can be independent cinemas or chains, negotiate with the distributor of the film they want to screen in the movie theatre. If the distributor has more than one version, that is, another version in addition to the original one, they will be able to choose their preferred version. Exhibitors typically choose only one version, although they could offer different ones, depending on the time, the day, or whether they can offer a different screen if it is a multiplex cinema. For instance, one cinema could offer the original version with subtitles in the country's language and the dubbed version of the same film to capture different kinds of demand. Also, in bilingual local

¹ The language to which she or he feels emotional attachment.

markets the versions in the two languages could also be offered (e.g. Québec in Canada or Catalonia in Spain). Therefore, in the first step the distributor with the rights of distribution of the film decides to incur the cost of dubbing in a certain language if it wants to have an additional language version, as well as the original. In multilingual markets, the distributor can decide to incur the fixed cost of dubbing for each language version it wants to offer (for instance, a distributor in Spain of a film whose original language is English has to decide whether to incur the cost of dubbing for each language version: Spanish, Catalan, Basque and Galician²). In the second stage, exhibitors negotiate (non-exclusive) contracts with the distributor and decide which versions will be screened.

The aim of this paper is to analyse whether the structure of the industry affects the language diversity in the movies market, a cultural market in which language matters. More precisely, I study how the concentration at the distribution and the exhibition levels affects the supply of films in the local language in the bilingual region of Catalonia (Spain). By using web-scraping techniques, I built a unique data set of all the screenings in the cinemas of Catalonia during one year, which I merged with other sources of data: census data, information on the distributors of the *Institut Català d'Empreses Culturals* (Catalan Institute of Cultural Enterprises; ICEC) and geographical information of the *Institut Cartogràfic i Geològic de Catalunya* (Cartographic and Geological Institute of Catalonia, ICGC). Next, I conducted an empirical analysis to explore the effect of the three relevant factors: concentration at the distribution and the exhibition levels, and the demand for films in Catalan language. I found that the concentration at the distribution level reduces the percentage of films in Catalan by 4.04 percentage points compared with the counterfactual of perfect competition. The effect of the concentration at the exhibition level is not significant. This implies that without such market failure the total supply of films in Catalan would be 96% greater (8.24% instead of the actual 4.2%). Then, the effect was disentangled depending on the target audience of the film: children audience or general adult public. The results from the heterogeneous analysis by type of audience indicate that children have higher preference intensity over the language because the market is more responsive to them; the concentration at the exhibition level matters when it comes to this type of consumer. Moreover, the effects are robust to many alternative specifications.

Thus, this paper contributes to the scarce literature on language and competition (Doh-Shin et al., 2021), but it also relates to the literature on the movie-theatre industry (Leung et al., 2020; Orbach & Einav, 2007). In a broader scope, it also contributes to the literature on competition and vertical relations (Allain et al., 2016; Fauli-Oller & Sandonis, 2016; Gans, 2007).

The rest of the paper is organized as follows: in the next section I will provide some background to the market and the case study. In Section 3, the theoretical framework will be formulated. Afterwards, I will show the data and descriptive statistics in Section 4. The empirical methodology will be presented in Section 5 and will be followed, in Section 6, by the results, which provide empirical evidence and validate the main hypothesis. Finally, I will conclude with a brief summary of the results, as well as some of their policy implications .

² Spanish is the official language of the country. Catalan is the co-official language in three regions which represent 29.5% of the population, Basque in two regions, representing 6.1%, and Galician in one region representing 5.7% of the total population (source: *Instituto Nacional de Estadística*, 2021).

2. Background

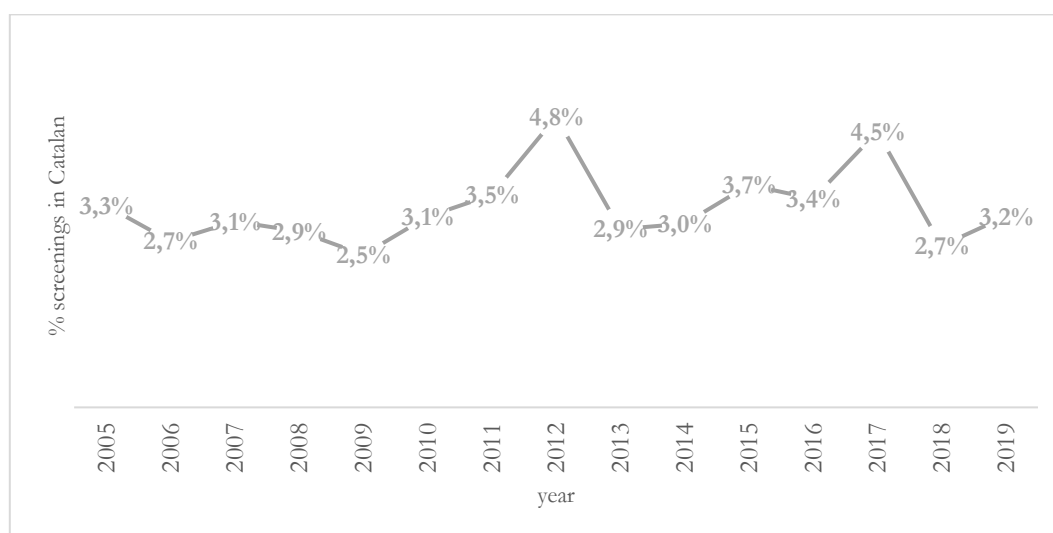
In this section I will present some stylized facts to better understand the movies market in the Spanish region of Catalonia.

The population of this Autonomous Community³ is above 7.5 million inhabitants⁴ (source: IDESCAT). According to the Survey on Language Uses of the Population of 2018, Catalan is the most used language for 36.1% of the population, 48.6% use mostly Spanish and 7.4% both (and the remainder use other languages). With regard to knowledge of the languages, 81.2% of the population are able to speak Catalan and 99.5% (virtually everyone) are able to speak Spanish. This indeed shows an asymmetry in knowledge, which corresponds with the framework that will be presented in the next section.

Thus, the case analysed is a good example of a bilingual market. In such markets we can find the two versions (of the two official languages) sometimes offered, either as dubbed films of foreign productions and the original version. In the last 15 years, there has been a low but stable supply of films in Catalan (see Figure 1).

Figure 1

Share of films in Catalan



Note: films in Catalan over the total number of screenings. Source: IDESCAT.

Moreover, among these films, except in 2010, the majority of them did not have Catalan as the original language. This means that most of these screenings are in Catalan due to the decision of the distributors and the exhibitors that chose to offer the Catalan version (see Table 1).

³ Administrative division of Spanish regions.

⁴ 7,716,760 according to IDESCAT in 2021. Provisional figure, from the Estimates of Population (IDESCAT) and “Cifras de población” (INE).

Table 1***Share of Films with Original in Catalan Out Of the Total Films in Catalan***

Note: The percentage is with respect to all the screenings in Catalan (dubbed, subtitled or original version).
Source: IDESCAT.

The supply of films in the Catalan version might seem disproportionately low if we compare it with other cultural goods such as TV, radio or theatre. The share of annual time watching TV in Catalan was around 20% from 2017 to 2020 (source: IDESCAT). If we look at the radio listeners, Catalan language broadcasters have more than 50% of the market (source: General Study of the Media, EGM; *Estudi General dels Mitjans*). Theatre performances are also mostly in Catalan, as it can be seen in the Table 2:

Table 2***Language in Theatres***

Theatre:	2015	2016	2017	2018	2019
Catalan	58.4%	64.4%	58.0%	59.9%	56.2%
Spanish	24.3%	21.0%	24.7%	23.7%	25.6%
Bilingual	8.2%	5.2%	7.0%	7.8%	7.7%
Non-spoken	7.1%	6.7%	7.3%	5.5%	7.8%

Source: IDESCAT

Indeed, these cultural markets are not perfectly comparable to the movies market in many aspects and, specifically, they can have a different demand (e.g. the average consumer of theatre is wealthier and more educated than the average consumer of cinema). However, such sharp differences cannot only be explained by the public that each cultural product attracts. Two things may happen: linguistic preferences change depending on the cultural good to be consumed, and/or the markets differ on the capability to match the demand for each language version.

The *Òmnibus* survey of the Centre of Opinion Studies (in Catalan, *Centre d'Estudis d'Opinió*) of the Catalan government asked about the language version preference over different cultural goods in several waves (2014, 2015, 2018 and 2019). If we look at the most

recent wave (2019), we see that for the same random sample of individuals, the preference regarding the language version changes depending on the cultural good (see Table 3).

Table 3

Preferences Over Language for Cultural Goods

Version	Books	Cinema	Theatre	Music
Spanish	47.42	49.08	36.92	17.42
Catalan	23	13.08	25.17	5.42
Original version	5.92	17.17	8.67	41.67
Indifferent	19.75	17.58	23.0	32.58
Other	2.42	1.08	0.42	2.33
Does not consume it	0.75	0.92	1.08	0.08
No answer	0.17	0.58	1.58	0.17
Does not know	0.58	0.5	3.17	0.33
Total	100	100	100	100

Source: Omnibus 2019, *Centre d'Estudis d'Opinió*

This table does not represent the language demand for each good, but the preferences of this random sample of 1,200 individuals. Thus, the variability between products is the result of the pure change of the revealed linguistic preference. For instance, if we restricted the sample to those who answered that they actually consume theatre periodically, the Catalan version is preferred by a 35.6% and the Spanish version is preferred by a 29.2%; that would explain the predominance of performances in Catalan in the theatre market.

Even though there is a disparity in the linguistic preferences over different cultural goods, that cannot explain by itself the differences in supply that were presented before. As we could see in Figure 1, the percentage of screenings in Catalan has been wavering around 3–4%. Throughout the waves of this *Omnibus* survey, the preference for cinema in Catalan has always been much higher than that: 21.7%, 24.2%, 20.9% (2014 waves), 22% (2015), 15.9% (2018) and 13.1% (2019). Hence, the supply of films in Catalan has always been between 5 and 7 times lower than these revealed preferences.

One concern regarding the responses of the survey might be that they are somehow politically motivated. If there was some political bias that, for instance, made indifferent people answer ‘Catalan’, such bias should be similar in the four cultural goods that are shown in the table. Even in this case, the disproportion between the actual supply and these declared (and perhaps biased) preferences would still be greater in the case of Cinema. On the other hand, the drivers of the language preference are irrelevant to the market: better knowledge of the language, emotional attachment or national identity, they all have the same effect in terms of demand behaviour: to be more inclined to one version rather than another.

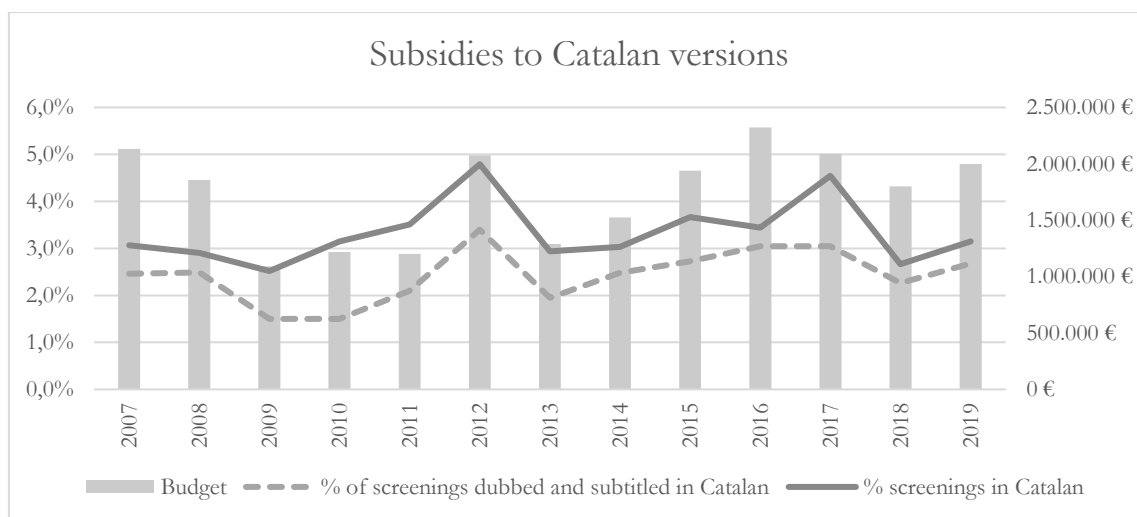
The debate about the undersupply of cinema in Catalan is not new. On 30th June in 2010, the Catalan Parliament approved by 117 votes out of 135 (87%) the so-called Law of the Cinema (in Catalan, *Llei del Cinema*), which established that 50% of the productions should be either dubbed or subtitled in Catalan. Later, the Constitutional Court declared that such a percentage was unconstitutional and lowered it to 25%. Then, with the quota of 25%

approved by the Constitutional Court, it was time to apply the law. Nonetheless, the Catalan government never applied it, perhaps due to the political pressure from the exhibitors and distributors, who threatened to stop supplying cinemas in Catalonia.

Since the regulation was not effective, the Catalan government has tried to influence the supply of films in Catalan through the dubbing costs, by subsidizing them.⁵ The budget efforts for such item have varied over time, depending on the budget constraints of the moment, as well as the political preferences of the party in power. On 27th July (2021), a major increase in this budget item was announced, reaching the peak of 2,705,000€. The historical series is presented in Figure 2:

Figure 2

Subsidies for Catalan Versions and Films in Catalan



Source: Language policy reports of the Language Policy Secretariat (Government of Catalonia).

Apparently, the subsidies do have an effect on the supply of screenings in Catalan. Nonetheless, precisely because the costs of dubbing are much lower with the subsidies, the gap between the declared demand (in the surveys) and the supply have to be explained by other factors.

The question is, therefore, if it is true that there is an undersupply of films in Catalan language, why does this happen? Is there any market failure that leads to that outcome?

3. Theoretical Framework

In this section I present an informal, standard, theoretical framework that will provide a better understanding of the empirical results. In particular, the impact of concentration on language diversity will be interpreted as a measure of the inefficiency caused by market power.

⁵ Distributors can apply for the subsidies whenever they want to dub a film. These subsidies are non-competitive.

In this representation of the movie-theatre industry, consumers have preferences regarding the content of the film as well as the language. I consider two languages: strong and weak. Native speakers of the strong language are monolingual (they cannot watch the film in the weak language). In contrast, native speakers of the weak language are bilingual (they can watch films in either language) but they prefer to watch movies in the weak language. Thus, in such a context a film is always released in the strong language and the question is whether the weak-language version is also provided.

Upstream Market

Distributors own the distribution rights of a number of films and choose whether or not to produce the weak-language version.

Downstream Markets

These are geographically segmented. Consumer prices are exogenously fixed (Orbach & Einav, 2007). Downstream firms, the exhibitors, supply films in one or several local markets. If the weak-language version exists, then the distributor bargains with each of the exhibitors, among other things, about its possible exhibition in the local market (adoption).

In the absence of informational and other types of frictions, we should expect the contract to maximize the joint payoffs of the distributor–exhibitor pairing (efficient bargaining). If the pair chooses to exhibit the film in both languages, then:

1. They incur a fixed cost
2. They steal some consumers from rivals (other pairs, distributors–exhibitors). That is, they attract consumers with strong preferences over language version that like the content of other films better (business-stealing effect).
3. They attract consumers who in the absence of the weak-language version would stay at home (market expansion effect).

The business-stealing effect increases with the degree of competition (fragmentation of the supply) in both the upstream and the downstream markets. In one of the extremes, if the exhibitor is a local monopolist, then the business-stealing effect is zero. The lower the market share of the exhibitor in the local market then the higher the temptation to steal business from rivals. In addition, both the business-stealing effect and the market expansion effects increase with the fraction of consumers with a preference for the weak language (this can be thought of as the demand for the weak language).

Prediction in the Downstream Market (Adoption)

The rate of adoption of weak-language versions increases with (1) the degree of competition in the local market, and (2) the fraction of consumers who prefer the weak language.

For the weak-language production decisions, the analysis is analogous. If the distributor produces the weak-language version then: it incurs a fixed cost, and generates business stealing and market expansion effects. All these effects will be filtered by negotiation with the exhibitors. Once again, the higher the degree of competition in distribution the stronger the business stealing effect and hence the higher the provision of weak-language versions.

In addition, competition in the upstream market causes a countervailing effect: higher competition in the upstream market reduces the bargaining power of the distributor, vis-à-vis exhibitors. Hence, if the distributor appropriates a lower share of the rents, then it will be less willing to produce more weak-language versions. Similarly, more competition in the downstream markets also increases the distributors' bargaining power (smaller and/or more fragmented exhibitors will have lower bargaining power with distributors), which enhances the incentives to produce weak-language versions.

Prediction in the Upstream Market (Production)

The rate of production of weak-language versions balances two effects of different signs: (1) it increases with the business-stealing effect and (2) it decreases with the bargaining power effect. As a result, the effect of market power in the upstream market on language diversity is, in principle, ambiguous.

Thus, the expected results are the following:

- **Higher competition (more fragmented market) at the distribution level will have an ambiguous effect on language diversity;**
- **Higher competition (more fragmented market) at the exhibition level will have a positive effect on language diversity;**
- **Higher demand (greater share of consumers that prefer the weak-language version) will have a positive effect on language diversity.**

Regarding the welfare analysis, let us ignore for the moment the market expansion effect. Will the equilibrium fraction of films in the weak language be excessive or insufficient from a total surplus point of view? According to our theoretical framework, the answer is ambiguous. However, it is important to emphasize that in the realistic case that fixed costs are relatively high, so that the equilibrium fraction of films with two linguistic versions is low, then the equilibrium level of linguistic version is inefficiently low.

In order to simplify the presentation of the welfare analysis, let us focus on the case in which upstream and downstream firms are integrated. Note that when firms are not integrated and the exhibitors share some of the surplus generated by the dubbed version, distributors' incentives to invest in linguistic diversity are further diminished. When the firm chooses whether or not to produce a second linguistic version, it compares the additional audience stolen from rivals with the fixed cost. Instead, the social planner internalizes the fact that the extra profits of the firm and the lost profits of the rivals cancel each other out. Thus, it only compares the increase in consumer surplus (speakers of the weak language get a better match) with the fixed cost. Hence, private incentives are insufficient or excessive, depending on the size of the business-stealing effect relative to the enhanced consumer surplus effect. The business-stealing effect positively depends on the intensity of language preferences as well as on the degree of competition. Thus, a higher preference intensity raises both the extra consumer surplus and the business-stealing effect. However, the business-stealing effect increases with the degree of competition, and goes to zero as we approach monopoly. Thus, in markets with a low degree of competition, the equilibrium level of linguistic diversity will tend to be inefficiently low.

To deal with the case of markets with intense competition we can refer to some previous research. In a similar model of monopolistic competition, Caminal (2010) showed that the

extra consumer surplus is lower (higher) than the business-stealing effect when the fixed cost of producing an additional linguistic version is relatively low (high), whereas the business-stealing effect is similar in both cases. The reason is simple. If the fixed cost is low, then most rivals will also offer the weak-language version. In this case, a firm that does not offer the weak-language version only attracts a small fraction of consumers with a preference for such a linguistic version. Thus, when the firm chooses to introduce the weak-language version, only a small fraction of consumers benefits from it. As a result, the private incentives exceed the public incentives and the equilibrium level of language diversity is excessive. On the contrary, if the fixed cost is relatively high and rivals only rarely offer the two linguistic versions, the extra consumer surplus generated by the additional version is large, private incentives are lower than social incentives, and therefore the equilibrium level of language diversity is inefficiently low. In the market we study in this paper, the fraction of versions in the weak language is small, which suggests that the latter case is the one that is empirically relevant.

Finally, note that when enhanced language diversity generates additional consumption (market expansion effect), private incentives are insufficient, as the firm cannot appropriate all the rents generated by the additional consumption.

Summarizing, if the market expansion effect is strong enough or if the intensity of competition is low, or if the fixed cost of producing a linguistic version is relatively high, then the equilibrium level of linguistic diversity is insufficiently low from a welfare point of view.

4. Data and Descriptive Statistics

For the empirical analysis of this paper, I combine several databases to answer the research question. First, I use a data set of screenings in all the movie theatres in Catalonia from July 2020 to May 2021,⁶ which contains 181,978 observations. This data set includes information about the weekday, the time of the screening, the week, the name of the movie, the cinema and, our relevant variable, the version, which can be in Catalan, in Spanish, Original version with Catalan subtitles and Original version with Spanish subtitles (when the original language is neither Catalan nor Spanish). The source of the database is the newspaper *El Periódico*, one of the most read newspapers in Catalonia, which publishes the billboard every week.⁷ To gather all this information, I have used ‘web-scraping’ techniques, which enabled collecting in a systematic manner all the information from the billboard during the gathering period. This makes my data set quite unique, since I constructed it.

I also used web scraping to obtain the film’s country of production, as well as its genre. In this case, the information was taken from FilmAffinity, a movie recommendation website that works as a huge movie database with more than 125,000 films listed. This information was matched with the main data set using the name of the film.

Similarly, I web-scraped International Movie Database (IMDb), the largest movie database, to obtain information on the original language of the film.

⁶ Although during this period movie theatres were not working at full capacity due to the COVID-19 restrictions (especially in 2020), that should only affect the number of observations, which is still very large, but not the rationale explained in Section 3.

⁷ Each Friday the billboard is updated, so ‘cinema weeks’ run from Friday to Thursday.

In addition, I also included cinema characteristics, such as the property (private, public or semi-public) and the geolocation through its address. I included information about the chain it belongs to in the case where it is not an independent cinema. This is crucial because in order to compute the competition measure of the exhibitors (cinemas), I had to take into account that they do not compete against cinemas of the same chain, but rather coordinate, avoiding cannibalization.

Another source of data was the Census 2011, which gave me information about the demand for Catalan versions for each cinema. Although the census does not ask about the linguistic preference in films, it asks about the ability to speak Catalan; therefore, the percentage of bilingual people (able to speak Catalan) can be a good proxy of the demand for the Catalan version. Since the information is given by census tract, which is also geolocated, I can impute the proxy for demand to each cinema, as explained below.

Finally, I obtained the information of the distributor of each film through the Catalan Institute of Cultural Enterprises (in Catalan: *Institut Català d'Empreses Culturals*) and ComScore. That information was matched to the main data set through the name of the film, and allowed me to compute the market share for each distributor in my data set.

4.1. Measures of Competition and Market Power

In order to analyse how market power and competition can affect the likelihood that a screening is in the Catalan version, I had to compute a proxy for the two levels of competition: the distribution and the exhibition.

The distribution is characterized by a business-to-business market. The marketplace is the whole country in which the distributor has the rights of the film, so in this case it would be Spain. The distributor can decide to offer the Catalan version or not, in the case where the original version is not in Catalan, and if so it would then be up to the exhibitors to offer this version in some of their screenings. The distributor, hence, incurs the fixed cost of dubbing, although this is often subsidized by the Catalan government. The bargaining power of the distributor will depend on their market share, and that is why I compute the share of the market for each distributor in the data set (market as number of screenings) and use this variable as proxy for market power.

There are 89 distributors in the sample. It is important to note that this is the market share by screenings, not by tickets sold. I could get information on the distributors of 180,032 of the 181,978 film screenings. Table 4 shows the screenings of each distributor in the sample as well as their share.

On the other hand, the exhibition is a business-to-consumer market, in which movie theatres can belong to a chain or can be an independent cinema. The marketplace, unlike the distributors, is not the whole country but its area of influence. We cannot consider that two cinemas, which are, for instance, more than 20 km⁸ from each other, compete between them, since they might share few or no potential consumers. We try several radii in order to check whether a 10 km radius is in fact an appropriate baseline. Moreover, it is also unlikely that

⁸ Davis (2006) finds that 'geographic markets consist of at most 15-mile circles around theatres, probably less'.

two neighbouring cinemas that offer different films will compete on the language version: the consumer that wants to watch film A and prefers the Catalan version over the Spanish one, will decide to go to another cinema if it offers the same film A in their preferred language version, but will not change their preference for watching film A because another film B is offered in Catalan.⁹ The assumption is that the preference for film goes before the preference for the language version, and so each film has its own demand. Furthermore, I also consider that film screenings compete in a concrete point in time: screenings in different weeks do not compete with each other. If *Star Wars: Episode 1* was available in one cinema, we cannot consider that it competes in language version with all the cinemas that offered the same film many years ago. The same applies for weeks for a simple reason: the billboard is updated every week so the consumer only knows the films offered in the following 6 days. The language version of a film screening in a certain week cannot be a factor of differentiation with another cinema offering the same film the following week because the consumer cannot know about that. Thus, the proxy of competitive pressure will be the number of same film screenings in the surrounding area that specific week, excluding those of cinemas in the same chain. This allows for great variability, unlike the proxy of market power of the distributor, which only depends on the distributor.

Table 4

Distributors

Name	Screenings	Share	Name	Screenings	Share
WBI	39,623	22.01%	EURODF	479	0.27%
UPI	20,871	11.59%	FESTIVAL	476	0.26%
ACONTRA	13,975	7.76%	BOSCO	457	0.25%
DIAMOND	12,858	7.14%	INDP	439	0.24%
SONY	10,697	5.94%	PAYCOM	385	0.21%
FILMAX	9,321	5.18%	RITA&LUCA	355	0.20%
DISNEY	7,104	3.95%	SYLDAVIA	329	0.18%
FLINS	6,426	3.57%	BEGIN AGAIN	327	0.18%
BTEAM	6,315	3.51%	PACKMAGIC	325	0.18%
DEAPLANETA	6,181	3.43%	AVENTURA	305	0.17%
eOne	5,022	2.79%	BENECE	300	0.17%
SELECTAVISION	3,869	2.15%	PPI	246	0.14%
AVALON	3,797	2.11%	39 ESC	185	0.10%
CAMEL	3,586	1.99%	SEGARRA	173	0.10%
TRI	3,537	1.96%	SHER	164	0.09%
ALFAPICT	3,437	1.91%	SPLENDOR	134	0.07%
VERCINE	2,770	1.54%	ATALANTE C	101	0.06%
ADSO	1,846	1.03%	BARTON	99	0.05%
VERTIGO	1,819	1.01%	CAPRICCI	99	0.05%
VERTICE	1,782	0.99%	FOX	70	0.04%
KARMA	1,594	0.89%	PARK CIRCUS	69	0.04%
SURTSEY	1,477	0.82%	REVERSO	62	0.03%
CINEMARAN	1,337	0.74%	SURFILMS	60	0.03%
WAND - AVAL	905	0.50%	EMON	55	0.03%
Wanda	794	0.44%	FLAMINGO	52	0.03%
FILMIN	772	0.43%	PREMIUM	37	0.02%
GOLEM	699	0.39%	PACK	36	0.02%

⁹ The same applies for Original Version with Subtitles. By default, the different versions are considered to play a role for a given film. This assumption is later relaxed.

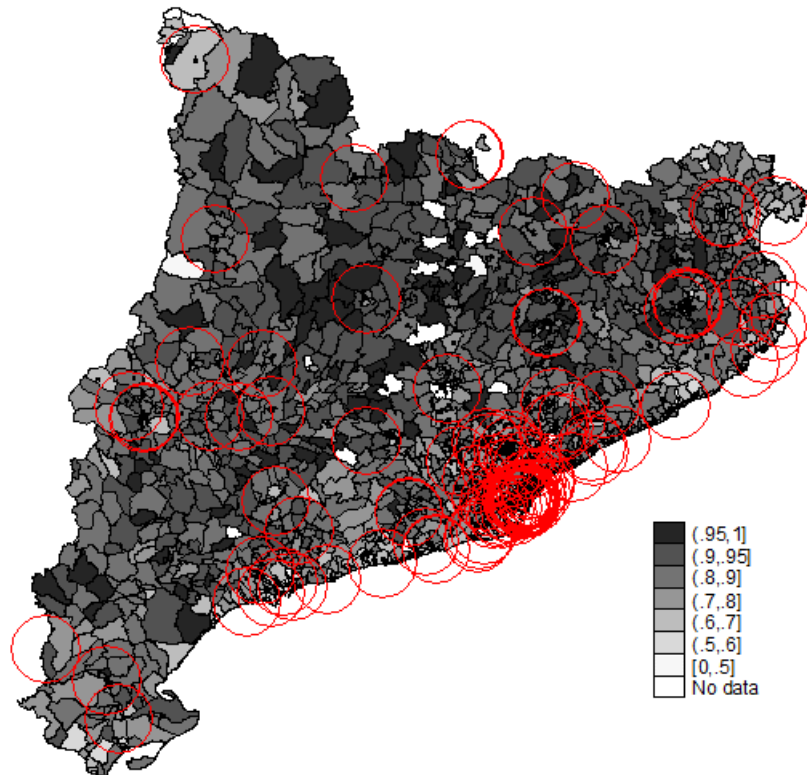
VERDIG	517	0.29%	MEDIA SOLUTIONS	33	0.02%
BIGPICTURE	501	0.28%	BARLOVENTO D.	25	0.01%
ELAMEDIA	492	0.27%	Others (30 distributors)	231	0.13%

4.2. Demand

It is impossible to know the precise demand¹⁰ for each language version since one individual could have different preferences for different films if they were bilingual: a spin-off film from a series that they used to watch in Catalan would lead to a preference for the Catalan version, even if, for the rest of the films, they had a preference for the Spanish version. However, since most of the films are not a spin-off, I consider that the main determinant will be the own language. The census data do not include information on the mother tongue or the identification language, but they give information on competences in Catalan (ability to understand, speak and write). In order to assign the demand to each cinema, I apply a similar method to the one I used when constructing the proxy of competitive pressure for exhibitors: I set a 10 km radius and assign all the census tracts whose centroids fell within the area of influence to the cinema. Hence, I can calculate the fraction of the population assigned to the cinema that is able to speak Catalan (see Figure 3).

Figure 3

Cinemas and Bilingual Population



¹⁰ Henceforth, 'demand' will mean 'the demand for the Catalan version'.

The average percentage of bilingual population in the data set (i.e. the 181,978 screenings) is 72.8%. This is lower than the percentage of the bilingual population (able to speak Catalan¹¹) of the Survey on Language Uses of the Population of 2018 (recall, 81.2%) because cinemas tend to be concentrated in urban areas, in which the percentage of bilingual population is below the mean.

4.3. Dependent Variable

Table 5 reports the language versions that we can find in the data set:

Table 5

Language Versions In the Dataset

Version	Freq.	Percent
3D Spanish Version	355	0.2
Catalan Version	7,666	4.21
Spanish Version	152,820	83.98
Original Version with Subtitles in Catalan	459	0.25
Original Version with Subtitles in Spanish	20,628	11.34
No information	50	0.03
Total	181,978	100

Only 50 screenings had no information on the language version (0.03%). Original versions with subtitles represent 11.59% of the total. Original versions with subtitles are not comparable to Spanish and Catalan versions, since the preference for the language of the subtitles might not be relevant for the individual, but rather the language of the original version (for which I do not have information). Putting together OVSC+CV and OVSS+SV¹² could be misleading. This is why I decided to not consider subtitled versions for the computation of the dependent variable. Therefore, the dependent variable *Catalan* will be a dummy that takes value 1 if CV and 0 if SV (or 3D SV).

4.4. Genres

From the FilmAffinity database I could web-scrape information about 169,778 screenings. In Table 6, I present the distribution of these screenings among the genres.

Table 6

Film Genres

Film Genre	Freq.	Percent	Cum.
Action	9,719	5.72	5.72
Animation	26,138	15.4	21.12
Adventures	142	0.08	21.2

¹¹ As explained in Section 2, virtually all individuals are able to speak Spanish.

¹² OVSC: Original Version with Subtitles in Catalan. CV: Catalan Version; OVSS: Original Versions with Subtitles in Spanish; SV: Spanish Version.

War	30	0.02	21.22
Science fiction	3,103	1.83	23.05
Comedy	27,409	16.14	39.19
Documental	7,239	4.26	43.46
Drama	36,162	21.3	64.76
Fantastic	6,264	3.69	68.45
Intrigue	3,619	2.13	70.58
Musical	4,113	2.42	73
Romance	8,828	5.2	78.2
TV Series	1,602	0.94	79.14
Terror	9,400	5.54	84.68
Thriller	25,572	15.06	99.74
Western	438	0.26	100
Total	169,778	100	

Note that 15.4% of the films are in the category Animation?. This distinction will be used later on in the analysis of the heterogeneous effects.

4.5. Summary Statistics

Table 7 reports the summary statistics of the relevant variables as well as the other control variables for the sample of the baseline model, that is, screenings which are either in Catalan or Spanish (not subtitled versions) and do have information on film genre.

Table 7

Summary Statistics

Variable	Mean	Std. Dev.	Obs
Catalan	0.042	0.201	152,707
ShareDist	0.095	0.077	152,707
CompExhib (10km)	42.78	71.923	152,707
CompExhib (20min drive)	48.839	79.017	152,707
Demand (10km)	0.732	0.051	152,707
Extension 1: films are substitutes within each genre			
ShareDist	0.269	0.262	152,707
CompExhib (10km)	109.008	165.505	152,707
Extension 2: upstream and downstream markets separately			
Catalan-Version	0.233	0.423	152,707
First stage, observation units are films			
Catalan-Version	0.149	0.356	626
ShareDist	0.044	0.060	626
Second stage, observation units are screenings but restricted to Catalan-Version=1			
Catalan	0.181	0.385	35,539
CompExhib (10km)	28.678	54.49	35,539
Demand (10km)	0.735	0.053	35,539
Control variables			
Public	0.005	0.072	152,707

Semi-public	0.016	0.124	152,707
Highly-educated (10km)	0.214	0.040	152,707
Unemployed (10km)	0.278	0.034	152,707
Born in Catalonia (10km)	0.643	0.047	152,707
USA origin	0.439	0.496	152,707
Spain origin	0.259	0.438	152,707
Monday	0.123	0.328	152,707
Tuesday	0.115	0.319	152,707
Wednesday	0.126	0.332	152,707
Thursday	0.127	0.333	152,707
Friday	0.148	0.355	152,707
Saturday	0.175	0.380	152,707
Sunday	0.187	0.390	152,707

The final sample consists of 152,707 screenings (the screenings for which I can merge all the variables and that do not have missing information). Only 4.2% are in Catalan, which is in the range of the supply of films in Catalan in the last 15 years (see Figure 1). We can see that the proportion of screenings of films that have the two versions is just 23.3% (the dummy Catalan-Version for the final sample). Among those films that do have the Catalan version available, 18.1% of their screenings are in fact exhibited in Catalan. This shows that there is room for an increase in language diversity both at the distribution and the exhibition level.

5. Empirical Methodology

The final purpose of this work involves analysing whether the market imperfections lead to a lower provision of film screenings in the Catalan version. The aim of the empirical analysis is to explain the probability of offering a screening in Catalan (vs Spanish) as a function of proxies for the demand and market power of the distribution and exhibition. The econometric specification, thus, relies on the following reduced form, Equation (1):

(1)

$$Catalan_{s,f,m,t} = \beta_0 + \beta_1 ShareDist_{d(f)} + \beta_2 CompExhib_{s,f,m,t} + \beta_3 Demand_m + \delta'W_f + \gamma'W_m + \theta_{c(m)} + \theta_t + \varepsilon_{s,f,m,t}$$

Here, $Catalan_{s,f,m,t}$ is a binary variable that takes 1 if the screening s of film f in the movie theatre m at time t is in the Catalan version, and 0 otherwise; we do not include subtitled screenings. There are three parameters of interest. The first one (β_1) corresponds to the variable $ShareDist_{d(f)}$, which is the proxy of the distributor's market power and it is at the distributor d level, which depends on film f . The second relevant coefficient (β_2) corresponds to $CompExhib_{s,f,m,t}$, which is the proxy of exhibitor competitive pressure, as explained in the previous section, which depends on screening s of film f in the movie theatre m at time t . Finally, the third relevant coefficient (β_3) corresponds to the $Demand_m$, which is the percentage of the Catalan-speaking population within the specified radius (10 km) around the movie theatre m .

I also include several control variables and fixed effects in the model. Specifically, W_f and W_m are vectors of control variables at the film level and movie theatre level, respectively. As for film level controls, I include country of production as well as genre, and by movie theatre level we include the property type (private, public or semi-public). Moreover, with the aim of not confounding the demand effect with the impact of other sociodemographic characteristics of the local consumers, I also control for the percentage of highly educated individuals, unemployment rate and percentage of individuals born in Catalonia at the local level (within 10 km), which are contextual factors that are likely to be correlated to the share of individuals who are able to speak Catalan. $\theta_{c(m)}$ are fixed effects at the chain c level, which depends on the movie theatre. The term and θ_t include time fixed effects: weekday and week. By including these fixed effects, I take into account the effect of price, which varies between chains and days, but it is fixed for a given cinema and day (all films at the same price).¹³ This econometric specification is based on the demand model of Davis (2006), which defines the utility of a consumer who watches film f in the movie theatre m ,¹⁴ plus the supply side proxies and controls.

This reduced form includes in the same equation the market power of the distributor (upstream market) and the competition at the exhibition level as well as the local demand for Catalan versions (downstream market). We also use two separate equations for the two stages as an alternative specification, as will be explained in Section 5.3.

Due to the nature of my dependent variable (dummy), I use a Linear Probability Model in order to get a direct interpretation of the coefficients.¹⁵ If the supply of films in the Catalan version was only driven by the demand, β_1 and β_2 should not be statistically different from 0, while β_3 should be positive and statistically significant. If market imperfection determined to some extent the supply of films in the Catalan version, as explained in Section 3, we should see that both β_1 and β_2 would be statistically different from 0. β_1 is in principle ambiguous, since the business-stealing effect might be offset by the bargaining effect. Nonetheless, we expect the former to be predominant, making the coefficient negative (the higher the market share, the lower the incentives to offer the Catalan version). β_2 , in turn, should be positive (the higher the number of competing screenings, the higher the incentive to compete on the language version).

The threats to the identification of β_1 and β_2 are due to unobserved (or not included) variables that can affect the probability that a screening is in the Catalan version and are correlated to $ShareDist_{d(f)}$ and $CompExhib_{s,f,m,t}$, respectively. The main threat to the identification of β_1 was that the original language of the movie was related to the distributor share of the market. Indeed, the so-called majors (big distributors) are often American and they distribute Hollywood movies in Spain, which tend to be blockbusters. We could potentially see that local distributors with a small share of the market tend to distribute films in which the original language is Catalan, while big distributors, instead, distribute foreign productions in which the original language is English. That would lead to an upward bias and β_1 could capture other things not directly related to the market power of the distributor.

¹³ I also tried a time-specific trend by chain and obtained very similar results. Available upon request.

¹⁴ In Davis (2006) this is referred to as h .

¹⁵ I also tried to estimate the same equation using non-linear binary choice models (probit and logit) and compute the corresponding average marginal effects. The results are reported in the Appendix.

For this reason, I include the W_f , which specifically consists of two dummies for Spanish origin and US origin; hence, we rule out the possibility that β_1 captures other effects.

Regarding β_2 , one of the main threats to its identification was the demand itself; indeed, without controlling for the proxy for the demand, there would be an omitted variable leading to a downward bias, since the demand and $CompExhib_{s,f,m,t}$ are spuriously correlated: the percentage of Catalan-speaking population is higher in rural areas, in which cinemas do not have competitors nearby. However, even controlling for demand-side effects and other sociodemographic controls at the local level, there could be other variables that, if not included, would lead to a downward bias, such as W_m , the property type of the cinema. Public and semi-public tend to offer a higher percentage of films in the Catalan version and they have a greater presence in less populated areas, in which there are no private cinemas even close by, and therefore the City Council tries to supply movies on its own (e.g. summer cinema outdoors).

Even if there could be some other unobserved factors that undermine the validity of the estimation, by including the time and chain fixed effects θ I make sure to control by all the unobserved heterogeneity related to time and chains. Moreover, since the price is fixed for a given cinema and day (it does not vary between films), the fixed effects will also capture the effect of the price.

Last but not least, standard errors are clustered at the cinema level since it is the level of variation of most of the aggregated variables of interest.

5.1. Heterogeneous Effects

After the analysis of the effect of the distribution and exhibition concentration on the supply of films in the Catalan version, using the data on the film genre, I can analyse the existing heterogeneous effects, that is, the effects of two different types of consumers: the children and the adults.

These two groups of consumers might differ on the linguistic preference for two reasons: there are no monolingual children since they learn both languages in school; and the intensity of the preferences are likely to be greater. Although children do not choose by themselves the films they watch, we can consider parents' decisions to be a good representation of their child's preferences; when a parent chooses to go to the cinema with her/his child, he or she tries to maximize the utility of the child. If the intensity is greater, cinemas will have a higher incentive to compete on the language version to attract more consumers for those movies targeted at children, that is, both the market expansion and the business-stealing effects will be greater. That would be further evidence of the market failure explained in Section 3.

In order to test for this heterogeneous effect, I use the dummy $Animation_f$, that takes the value 1 if the film's genre is 'Animation' and 0 otherwise. I assume that the 'Animation' genre is for films targeted at children and the others target the general adult audience. $Animation_f$ is interacted with $ShareDist$, $CompExhib$ and $Demand$ to disentangle the effect of these variables for adults and children separately.

5.2. Robustness Checks

In order to test the robustness of the results, several sensitivity checks are performed, both to the baseline model and to the model allowing for heterogeneous effects by the genre of the film.

As explained in Section 4.1., to check whether 10 km is an appropriate radius, we try several radii, both together and separately (5 km, the default 10 km, 15 km and 20 km). According to Davis (2006), ‘geographic markets consist of at most 15-mile circles around theatres, probably less’. Also, we should keep in mind that the analysis was made in the US, which is more car-based than European countries, and thus such local markets might likely be smaller.

In addition to that, I also use a different definition of the relevant variables *CompExhib* and *Demand*. Recall that the boundaries of the local market were defined as a radius of 10 km from the location of the cinema, thus I could count the number of competing films in the same week (*CompExhib*) and the proportion of bilingual speakers among the population older than 2 years (*Demand*). Alternatively, I set the boundaries using a measure of accessibility, the driving time. More specifically, I set the boundaries of the local market to a 20-minute driving distance from the cinema. I computed this with the command ‘georoutei’ in Stata, which uses the Web mapping HERE WeGo. This new measure implicitly takes into account the geography, connections, as well as the rural/urban difference. Thus, *CompExhib* and *Demand* are recalculated.

Furthermore, since a ‘preference towards the original’ exists, and although it would be possible from a theoretical point of view, it is not a common practice to offer another version rather than the original for those films whose original language is Spanish or Catalan. Except for animation films, because they are not real actors and therefore do not have an original voice, I restrict the sample to only those films whose original language is a foreign language, and therefore in which no ‘preference towards the original’ might bias the results.

5.3. Extensions

The model relies on a strong, although realistic, assumption: each film has its own demand, they are not substitutes for each other (the language version cannot shift demand from one film to another). The extension I propose, which represents an additional robustness check, consists of assuming that similar films are substitutes. More precisely, films of the same genre are considered to be perfect substitutes. Hence, animation films will be competing in the same market, as well as drama, comedy and so on (see Table 6, all the genres in the data set).

The two competition measures at the two steps – distribution and competition – are redefined accordingly. That is, the share of the market of the distributor (*ShareDist*) is (re)computed as the market share within each ‘genre’, rather than for the whole market. If most of the screenings of animated films belong to a certain distributor, such a distributor will have a very high share even if it is non-present in other genres. Moreover, the number

of competing films (*CompExhib*) is obtained considering all the films of the same genre offered by nearby competing exhibitors during that week.

The second extension, which also works as an additional robustness check, consists of separating the two stages in two equations. Thus, in the first step I look at the probability that a distributor offers the Catalan version of a film, therefore films are the observation units; and in the second step, I look at the probability that a screening is offered in Catalan in a movie theatre, restricting the sample to those films that do have a dubbed version (so at the first stage the distributor incurred the cost of dubbing in Catalan).

First step¹⁶:

$$Catalan_Version_f = \beta_0 + \beta_1 ShareDist_{d(f)} + \delta'W_f + \theta_t + \varepsilon_{s,f,m,t} \quad (2)$$

Second step:

$$Catalan_{s,f,m,t} = \beta_0 + \beta_2 CompExhib_{s,f,m,t} + \beta_3 Demand_m + \delta'W_f + \gamma'W_m + \theta_{c(m)} + \theta_t + \varepsilon_{s,f,m,t} \quad \text{if } Catalan_Version_f = 1 \quad (3)$$

6. Results

The results of the baseline model are reported in Table 8. In column (1) we observe that indeed there is a negative and significant effect of the market power of the distributor in the language diversity, but the coefficient of *CompExhib*, the β_2 , is negative and significant. This is reversed once I include the *Demand* in column (2). Local markets with a higher degree of competition (urban areas) also have a lower percentage of bilingual speakers. By including *Demand*, such spurious correlation is driven out and the coefficient β_2 becomes positive, although not significant. In column (3) the controls at the film level, W_f (dummies of genre and country of origin), are included. β_2 leaps up in magnitude and significance, although remains non-significant even at 10%; β_3 remains stable, and β_1 increases its magnitude and significance. In column (4) the controls at the movie-theatre level, W_m (property type and sociodemographic characteristics of the local market), are included. The main change here is the increase of β_3 , since previously it was downward biased by the correlation with local sociodemographic characteristics. Finally, the full specification is used in column (5) by including the fixed effects $\theta_{c(m)}$ (ownership dummies) and θ_t (weekday and week dummies); the coefficients remain quite stable.

¹⁶ θ_t are dummies for weeks since the initial period, not days. A film can be on screen for more than one week.

Table 8

LPM, Baseline Model

	(1)	(2)	(3)	(4)	(5)
	Catalan	Catalan	Catalan	Catalan	Catalan
<i>ShareDist</i>	-0.323*** (-4.88)	-0.338*** (-4.99)	-0.476*** (-6.15)	-0.466*** (-5.97)	-0.425*** (-5.53)
<i>CompExhib</i>	-0.0121** (-3.15)	0.00259 (0.75)	0.00534 (1.53)	0.00575 (1.32)	0.00686 (1.49)
<i>Demand</i>		0.506*** (3.38)	0.506*** (3.74)	0.632** (2.91)	0.613** (2.90)
<i>US origin</i>			-0.0502*** (-4.57)	-0.0499*** (-4.50)	-0.0493*** (-4.43)
<i>Spain origin</i>			-0.00102 (-0.12)	-0.000275 (-0.03)	-0.00439 (-0.54)
<i>Public</i>				0.0416 (1.65)	0.0424 (1.54)
<i>Semi-public</i>				0.0914 (1.39)	0.0984 (1.58)
<i>Highb-educ</i>				-0.499** (-2.97)	-0.305 (-1.50)
<i>Unemployed</i>				-0.664*** (-3.99)	-0.336 (-1.54)
<i>Cat-born</i>				-0.501* (-2.17)	-0.441+ (-1.88)
<i>Genre dummies</i>	No	No	Yes	Yes	Yes
$\theta_{c(m)}$	No	No	No	No	Yes
θ_t	No	No	No	No	Yes
<i>Const</i>	0.0779*** (5.83)	-0.297** (-2.83)	-0.341*** (-3.40)	0.171 (1.22)	-0.0078 (-0.06)
N	152707	152707	152707	152707	152707

t statistics in parentheses + p<0.1, * p<0.05, ** p<0.01, *** p<0.001
 Clustered standard errors at the cinema level. CompExhib rescaled by 100.

Considering the full specification displayed in column (5) of Table 8, it appears that the market power of the distributor has a strong and significant negative effect on the supply of films in Catalan. For a 1 percentage point (p.p.) increase in the share of the distributor, the probability of supplying a film in Catalan will decrease by a 0.425 p.p.

In this specification, the competition at the exhibition level does not seem to have a significant effect. However, the results presented in the next subsection indicate that it indeed matters, depending on the target audience.

In order to get a better idea of the scope of the effect of the competition on the language version, we can compute a counterfactual. The monopolistic behaviour of the distributor is due to its market power. In fact, small distributors, with a share of the market close to zero, do not have such monopolistic behaviour. Thus, we can compute the total effect of the monopolistic behaviour at the distribution level by assuming a mean of *ShareDist* equal to 0 (hypothetical perfect competition):

$$\beta_1 * \text{Mean}(\text{ShareDist}_{d(f)}) - \beta_1 * 0 = 0.425 * 0.095 = 4.04 \text{ p. p.}$$

This means that the concentration at the distribution level leads to an undersupply of 4.04 p.p. of films in the Catalan version. Such a figure might be low at first sight, but considering that the percentage of films in the Catalan version in our sample is 4.2%, that implies that without such monopolistic behaviour the share of films in Catalan would be 96% greater (8.24% instead of the actual 4.2%).

With regard to the *Demand*, the interpretation of the coefficient β_3 is also simple: for a 1 p.p. increase of bilingual speakers in the local market, the propensity to offer a film in the Catalan version would be 0.613 p.p. higher.

6.1. Heterogeneous Effects

Table 9 reports the results of the analysis of heterogeneous effects. In the first column, the aggregated effects are displayed; they are the coefficients of the three relevant variables of the full specification of the baseline model (column (5) of Table 8). In column (2) we can see the coefficients for adults and children, respectively. As explained in Section 5.1., we interact *ShareDist*, *CompExhib* and *Demand* with the dummy *Animation*, which indicates whether it was an animated film or not, assuming that animated films are targeted at children. In column (3) the equality tests are shown; they are all significant, meaning that $\beta^{children} \neq \beta^{adults}$ for none of the three relevant variables. As can be seen, the magnitudes are greater for the $\beta^{children}$. This should be interpreted as a clear sign of differentiated behaviour of firms (distributors and exhibitors), depending on the type of audience.

It is also important to note that $\beta_2^{children}$ is positive and significant. An increase in the competition at the exhibition level leads to an increase in language diversity but only for animated films. More precisely, for an increase of 100 competing screenings, the supply of movies in Catalan will increase by 4.06 p.p.

Table 9

Heterogeneous Effects

	(1) Aggregated effect	(2) Heterogeneous	(3) Equality test (p-value)
<i>ShareDist</i>	-0.425*** (-5.53)		
<i>CompExhib</i>	0.00686 (1.49)		
<i>Demand</i>	0.613** (2.90)		
<i>ShareDist - adults</i>		-0.1835*** (-4.29)	
<i>ShareDist - children</i>		-1.157*** (-6.69)	0.0000
<i>CompExhib - adults</i>		-0.0012 (-0.69)	
<i>CompExhib - children</i>		0.0406*** (4.14)	0.0001
<i>Demand - adults</i>		0.341+ (1.73)	
<i>Demand - children</i>		1.864*** (5.65)	0.0000
N	152707	152707	

t statistics in parentheses + p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Clustered standard errors at the cinema level. CompExhib rescaled by 100. The aggregated effect refers to the full specification of the baseline model, column (5) of Table 8.

The market is clearly more sensitive to the child audience than the adult, suggesting that indeed children have higher intensity of preferences.¹⁷ This, in turn, means that the undersupply of a Catalan version is greater for child-targeted movies.

6.2. Robustness Checks

The decision on which radius should be applied to define the local market is based on Davis (2006), although he applies a shorter distance. To check whether the decision of 10 km was appropriate, I used different radii for the definition of the variable *CompExhib*: 5 km, 10 km (the default), 15 km and 20 km: first separately, and then together in two variables *CompExhib*_{10km} and *CompExhib*_{10km-20km} (the number of competing screenings from 10 km to

¹⁷ An alternative explanation could be that the parents have a higher intensity of preferences when choosing a film for their children rather than for themselves.

20 km away). Since the aggregate effect of *CompExhib* is not significant for none of the radii, Table 10 reports only the coefficients of children. Since the variable *Demand* is also affected by the change in radius, it is also reported.

Table 10

Different Radii

	(1)	(2)	(3)	(4)	(5)
<i>CompExhib – children</i> <i>5km</i>	0.0596** (2.89)				
<i>CompExhib – children</i> <i>10km</i>		0.0406*** (4.14)			0.0382** (2.73)
<i>CompExhib – children</i> <i>15km</i>			0.0293*** (3.77)		
<i>CompExhib – children</i> <i>20km</i>				0.0256*** (4.47)	
<i>CompExhib – children</i> <i>10km - 20km</i>					0.0126 (0.90)
<i>Demand – children</i> <i>5km</i>	1.3923*** (5.89)				
<i>Demand – children</i> <i>10km</i>		1.8644*** (5.65)			2.1157*** (4.89)
<i>Demand – children</i> <i>15km</i>			1.7743*** (4.41)		
<i>Demand – children</i> <i>20km</i>				1.7669*** (4.28)	
<i>Demand – children</i> <i>10km - 20km</i>					0.5951 (1.10)
N	152707	152707	152707	152707	152707

t statistics in parentheses + p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Clustered standard errors at the cinema level. *CompExhib* rescaled by 100. The radii of the control variables *High-educ*, *Unemployed* and *Cat-born* are also adapted in each equation. In equation (5) they are included with radius 10km and from 10 to 20km, altogether.

In column (1) to (4) we can see that there is a decay of β_2 as we increase the size of the local market, while the *Demand* coefficient β_3 remains substantially stable. Once we look separately at the effect of the competition within 10 km distance and between 10 km and 20 km distance, we see that the competition in the outer circle does not have any effect, and the same applies for the *Demand*. This result is relevant for two main reasons. First, because it supports the decision of choosing 10 km as the baseline radius for the local market. Second, it shows that, unlike Davis (2006), the competition does not matter beyond 10 km; according

to Davis (2006), ‘geographic markets consist of at most 15-mile circles [24 km approx.] around theaters, probably less’. We find an even stricter limit of such geographic markets. This might be due to the fact that the case study of Davis was conducted in the US, a more car-based country in which individuals are used to driving longer distances compared to European countries.

In Table 11, I present the two additional robustness checks for the aggregated effects, column (1) and (3), and for the heterogeneous effects, column (2) and (4). The restriction of the sample to films whose original language is neither Catalan nor Spanish (the two official languages in Catalonia) aims to avoid a ‘preference towards the original’ bias. In this case, hence, we make sure that the preference over the language version has nothing to do with other characteristics that might be correlated with the relevant variables. The results are very robust; β_2 is actually much higher and significant for the aggregated effect in column (1). The results of the heterogeneous effects in column (2) show that the coefficients are also very robust here. The equality tests are highly significant, meaning that $\beta^{children} \neq \beta^{adults}$ for the three relevant variables.

The redefinition of the boundaries, from 10 km to 20-minute driving distance, does not lead to different conclusions. The direction and significance of the coefficients in both column (3), for the aggregated effects, and column (4) for the heterogeneous effects, are the same as those we found using the 10 km radius. The magnitudes of the coefficients do not vary considerably. This is strong and robust evidence in favour of the results of Table 8 and Table 9.

Table 11***Robustness Checks***

	(1) Foreign Languages	(2) Foreign Languages	(3) Redefined boundaries (20min drive)	(4) Redefined boundaries (20min drive)
<i>ShareDist</i>	-0.462*** (-5.14)		-0.425*** (-5.51)	
<i>CompExhib</i>	0.01686** (2.64)		0.00634 (1.54)	
<i>Demand</i>	0.702** (3.05)		0.452** (2.60)	
<i>ShareDist - adults</i>		-0.0999+ (-1.83)		-0.185*** (-4.29)
<i>ShareDist - children</i>		-1.119*** (-7.25)		-1.168*** (-6.75)
<i>CompExhib - adults</i>		0.0064 (1.15)		-0.00051 (-0.15)
<i>CompExhib - children</i>		0.0432*** (3.90)		0.0392*** (4.13)
<i>Demand - adults</i>		0.3010 (1.39)		0.2008 (1.33)
<i>Demand - children</i>		1.8408*** (5.54)		1.7376*** (5.05)
N	95514	95514	152707	152707
Equality tests (p-values)				
<i>ShareDist</i>		0.000		0.000
<i>CompExhib</i>		0.001		0.000
<i>Demand</i>		0.000		0.000

t statistics in parentheses + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Clustered standard errors at the cinema level. CompExhib rescaled by 100.

6.3. Extension

In this subsection the results of the two extensions explained in Section 5.2. are presented.

Table 12:

Extension 1. Films Are Substitutes if Same Genre

Extension 1	(1) Aggregated effect	(2) Heterogeneous	(3) Equality test (p-value)
<i>ShareDist</i>	-0.168*** (-5.79)		
<i>CompExhib</i>	0.0071*** (4.28)		
<i>Demand</i>	0.640** (3.07)		
<i>ShareDist - adults</i>		-0.0886*** (-4.44)	
<i>ShareDist - children</i>		-0.8877*** (-6.37)	0.0000
<i>CompExhib - adults</i>		0.0043** (2.96)	
<i>CompExhib - children</i>		0.0250** (2.84)	0.0181
<i>Demand - adults</i>		0.368+ (1.90)	
<i>Demand - children</i>		1.818*** (5.64)	0.0000
N	152707	152707	

t statistics in parentheses + p<0.1, * p<0.05, ** p<0.01, *** p<0.001
 Clustered standard errors at the cinema level. CompExhib rescaled by 100.

Table 12 reports the results of the first extension. Here, *ShareDist* and *CompExhib* are recalculated, taking into account that films compete within each genre. The magnitude of β_1 and β_2 is indeed different, but the direction and significance of the coefficient support the hypothesis of the paper and prove the robustness of the previous results. The main difference in this new extension is that the competition at the exhibition level does not only affect the language diversity in animated films, but in all kinds of films. Thus, the previous results (in Tables 8 and 9) might provide lower bound estimates, since in this formulation adult-targeted movies are also affected. The equality tests are significant for the three relevant variables, indicating that indeed the market is more sensitive to children than adults: $|\beta^{children}| > |\beta^{adults}|$.

Finally, I present in Table 13 the results of the two stages separately. As explained in Section 5.3., in the upstream market the distributor has to decide whether to incur the cost of dubbing to offer the Catalan version or not, for each film (not screening, this is why the film is the observation unit). In the downstream market, the exhibitors (the cinemas) have to

choose whether to offer the version in Catalan or Spanish, for those films in which a Catalan version is available (and therefore the sample is restricted to screenings of films with a Catalan version available).

Let us first focus on the aggregated effects, in column (1) and (3). We can note that the coefficients are equal in significance and direction to the baseline model: a negative and significant effect of *ShareDist*, a positive and non-significant effect of *CompExhib*, and a positive and significant effect of *Demand*. Nonetheless, the effect of the competition at the exhibition level is again significant when we separate animated films from the rest, as we observed in Table 9. Both $\beta_2^{children}$ and $\beta_3^{children}$ are significantly higher than their adult counterparts, according to the equality tests. The lack of significance of $\beta_1^{children}$ and its equality tests is mainly due to the small size of the sample used (recall: films, not screenings). Among the 626 films of this sample, only 77 are animated films. These are too few observations to lead to significant results both in the $\beta_1^{children}$ and the corresponding equality test. However, if we look at the point estimate, we can note that $\beta_1^{children}$ is indeed higher than the β_1^{adults} . Overall, these second extension's results, reported in Table 13, provide evidence that the results found in Table 8 and Table 9 are consistent.

Table 13

Extension 2. Two Stages Separately

Extension 2	Upstream market (1 st stage)		Downstream market (2 nd stage)	
	(1) Catalan Version	(2) Catalan Version	(3) Catalan	(4) Catalan
<i>ShareDist</i>	-0.605** (-2.94)			
<i>CompExhib</i>			0.0145 (1.03)	
<i>Demand</i>			2.043** (2.57)	
<i>ShareDist - adults</i>		-0.5225** (-2.66)		
<i>ShareDist - children</i>		-1.552 (-1.24)		
<i>CompExhib - adults</i>				-0.0136 (-0.65)
<i>CompExhib - children</i>				0.0283* (2.08)
<i>Demand - adults</i>				1.554+ (1.92)
<i>Demand - children</i>				2.463** (3.09)
N	626	626	35539	35539
Equality tests (p-values)				
<i>ShareDist</i>		0.416		
<i>CompExhib</i>				0.022
<i>Demand</i>				0.000

t statistics in parentheses + p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Clustered standard errors at the cinema level. CompExhib rescaled by 100. In the upstream market we consider films as the observation units. In the second stage, we restrict to those screenings which have a Catalan version.

7. Conclusion

This paper represents, to the best of my knowledge, the first attempt to empirically analyse the language supply of a cultural market. The aim of this work was to find whether the concentration of the movie-theatre industry, both at the distribution and the exhibition levels, led to an undersupply of language diversity in a bilingual context.

To do so, I built a unique data set, using advanced web-scraping techniques and combining different sources, of all the cinema screenings in the bilingual region of Catalonia to explore a potential market failure. Such failure would consist in an undersupply of films in the Catalan version due to the concentration in two stages of the market. Applying a Linear Probability Model controlling by relevant factors and adding several fixed effects, I found that the concentration of the market indeed leads to an undersupply of 4.04 percentage points of films in Catalan, which means that if those market failures did not exist there would be 96% more screenings in Catalan. This undersupply can be entirely attributed to the monopolistic behaviour of the distributors.

By using the genre of the film, I disentangle two types of consumers, children and adults. The market response differs if the film is targeted at children or at the general adult audience. Due to the higher intensity of the linguistic preference of the children compared to the adults, I found that the competitive pressure of the exhibitors only changes the supply of films in Catalan when it deals with the child demand (animation films), and the effect is negligible for the rest. These results were robust to several checks.

According to the standard theoretical framework I provide, the empirical evidence indicates that the movie-theatre industry underprovides language diversity. More specifically, the fraction of movies dubbed into Catalan is insufficiently low from a welfare point of view. Most of the inefficiency can be attributed to the upstream firms, the distributors, who act as a bottleneck. Such underprovision is higher in local markets with low competition when it comes to animated films, that is, films targeted at children; this additional effect can be attributed to the downstream firms. Therefore, a first policy implication that can be derived from these findings is that the focus should be put on the distributors, who in the first stage decide whether to incur the fixed cost of dubbing in order to offer an additional version in Catalan. Thus, subsidizing the costs of dubbing or establishing quotas of films distributed in Catalan would be policies that can be justified from an efficiency point of view. Also, another policy implication that can be derived from the findings on the heterogeneous effects is that the welfare loss due to the low provision of films in Catalan is greater for animated films, so policies could be more tailored and focus on child-targeted films. For instance, exhibitors could be legally bound to offer the Catalan version at least in one of the screenings of an animated film.

The finding on the heterogeneous effects sheds light on another important issue: the endogeneity of the preferences. Indeed, if Catalan-speaking adults accommodate their preference more than children (lower intensity of preference), it indicates that the linguistic preference changes over time, at least in intensity. The drivers of this change are not the focus of the research of this paper, but it would be an interesting topic for further research on linguistic preference in cultural markets.

References

- Allain, M-L, Chambolle, C., & Rey, P. (2016). Vertical integration as a source of hold-up. *The Review of Economic Studies*, 83(1) 1–25.
- Caminal, R. (2010). Markets and linguistic diversity. *Journal of Economic Behavior and Organization*, Elsevier, 76(3),774.
- Caminal, R., & Di Paolo, A. (2019) Your language or mine? The noncommunicative benefits of language skills. *Economic Inquiry*, 57(1), 726–750.
- Chen, Y., Riordan, M., 2007. Price and variety in the spokes model. *Economic Journal* 117, 897-921.
- Davis, P. (2006). Spatial competition in retail markets: Movie theaters. *The RAND Journal of Economics*, 37(4) 964–982.
- Doh-Shin J., Jullien, B., & Klimenko, M. (July 2021). Language, Internet and platform competition. *Journal of International Economics*, 131(103439).
- Fauli-Oller, R., & Sandonis, J (2016). Welfare effects of downstream mergers and upstream market concentration. *The Singapore Economic Review*, 61(05), 1550056.
- Gans, J. S. (2007). Concentration-based merger tests and vertical market structure. *The Journal of Law and Economics*, 50(4), 661–681.
- Ginsburgh, V., & Weber, S. (2011). *How many languages do we need? The economics of linguistic diversity*. Princeton University Press.
- Grin, F. (1992). Towards a threshold theory of minority language survival. *Kyklos*, 45, 66–97.
- Leung, T. C., Qi, S., & Yuan, J. (2020). Movie industry demand and theater availability. *Review of Industrial Organization*, 56, 489–513.
- Mèlitz, J. (2012). A framework for analyzing language and welfare. SIREDP-2012-89.
- Tucker, G. R. (2001). A global perspective on bilingualism and bilingual education. In J. E. Alatis, & A.-H. Tan (Eds.), *Roundtable on language and linguistics*.
- Orbach, B. Y., & Einav, L. (2007). Uniform prices for differentiated goods: The case of the movie-theater industry. *International Review of Law and Economics*, 27(2), 129-153.
- Wickstrom, B. A. (2005). Can bilingualism be dynamically stable? A simple model of language choice. *Rationality and Society*, 17(1), 81-115.

Appendix

Table A1:

Logit and Probit (Average Marginal Effects)

	(1) Logit	(2) Logit	(3) Probit	(4) Probit
<i>ShareDist</i>	-0.474*** (-7.33)		-0.4593*** (-7.45)	
<i>CompExhib</i>	-0.00052 (-0.13)		-0.00121 (-0.29)	
<i>Demand</i>	0.4247+ (1.95)		0.4615* (2.27)	
<i>ShareDist - adults</i>		-0.3885*** (-6.20)		-0.3924*** (-6.44)
<i>ShareDist - children</i>		-0.9115*** (-4.96)		-0.8853*** (-5.26)
<i>CompExhib - adults</i>		-0.0079** (2.66)		-0.00896** (-2.73)
<i>CompExhib - children</i>		0.0301*** (2.39)		0.0298** (2.63)
<i>Demand - adults</i>		0.220 (1.61)		0.2444+ (1.83)
<i>Demand - children</i>		1.425** (2.52)		1.5292** (3.10)
N	143042	143042	143042	143042
Equality tests (p-values)				
<i>ShareDist</i>		0.004		0.009
<i>CompExhib</i>		0.000		0.000
<i>Demand</i>		0.028		0.002

t statistics in parentheses + p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Clustered standard errors at the cinema level. CompExhib escaled by 100. Aggregated effects in columns (1) and (3). Heterogeneous effects in columns (2) and (4).

The logo for UBIREA, featuring the text 'UBIREA' in a bold, white, sans-serif font inside a white rounded rectangle.

UBIREA

Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

WEBSITE: www.ub.edu/irea • **CONTACT:** irea@ub.edu

The logo for AQR, featuring a green circle with a white dot inside, followed by the text 'AQR' in a bold, white, sans-serif font inside a white rounded rectangle.

AQR

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • **CONTACT:** aqr@ub.edu