Second-Language Acquisition Behavior and Hegemonic Language

Author. Kentaro Hatsumi
Address. Faculty of Policy Studies, Kansai University, Suita, Osaka, 564-8680, Japan
Email. hatsumi *atmark* kansai-u.ac.jp
This Version. November 26, 2021
(This version appears in International Journal of Economic Theory.)

Keywords. coalition formation game, coordination game, economics of languages, large game

JEL Classification Codes. C72, D85, F51, F68

Abstract. We construct a game theoretic model in which there are multiple countries with their own languages and each citizen can gain from additional communication in her secondarily acquired language. We demonstrate that in any equilibrium, a hegemonic language, which is a language that all citizens in other countries want to study, emerges. Such an equilibrium is more likely to exist if the size of the population of a country that is not the largest increases, or if the ratio of the gain from the additional communication in the second language to the cost for acquisition increases.

1 Introduction

Language plays a quite important role in various social aspects, including politics and economics. In political contexts, the classical and influential work by Anderson (1983) suggests that print media written in one language was critical in constituting a sense of unity of a modern nation state. In economic contexts, even within a country, differences in spoken language often work as entry barriers preventing language minorities from obtaining job opportunities.[1] It is worthwhile for social scientists to constitute models that focus on language and related issues and analyze them.

[1] Even in the United States or the United Kingdom, which are countries that are believed to have respect for diverse ethnicities and languages, empirical studies such as Tainer (1988), Dustmann and Fabbri (2003), Bleakley and Chin (2004) and Chiswick and Miller (2010) suggest that workers who have English proficiency can gain significantly higher wages. Budría et al. (2019) report that Spanish proficiency brings higher job opportunities for immigrants in Spain.

The choice of foreign languages to study by individuals and educational policy planners has been an important subject for a long time. [2] If one constitutes a model of second-language acquisition behavior, a standard approach may be the adoption of the assumption that language is a communication device and the number of people with whom one can communicate is important for her choice of second language.[3] In this paper, we take this standard approach to construct a model and derive some implications for understanding the linguistic situations in the world.

The model we construct is as follows. There are four countries.[4] This is for the sake of simplicity, and we can easily increase the number of countries.[5] Each country has its own language and a continuum of agents (or citizens). Each agent can speak her native language and can learn, at most, one foreign language. Each agent can gain linearly from the number of foreign agents with whom she can communicate in her acquired second language. Each agent has her own effort cost to learn a foreign language that is hidden at first. Even though this is a two-stage game for introducing the uncertainty of cost, it is essentially simultaneous. In the first stage, all agents choose their second language, and in the second stage, after the cost is revealed, they decide whether to actually study. We employ a relatively strong equilibrium concept, a partially strong subgame perfect Nash equilibrium, which prohibits all coalitional deviations within a country to study other languages in the first stage.

What we can obtain by analyzing this model is the following. In any equilibrium, there exists a hegemonic language, which is a language that all agents in other countries want to study (Proposition 2). Without any condition, an equilibrium exists in which the language of the country with the largest population is hegemonic (Proposition 3). For languages of the countries with the second, third, and fourth largest populations to become hegemonic, we need some conditions (Proposition 4). As the relative size of the population of the second, third, or fourth largest country to that of the largest country increases, any equilibrium in which the language of the second, third, or fourth largest country is hegemonic is more likely to exist

[2]It is very difficult to distinguish between "second" and "foreign" languages. Second language is a more useful term because second languages can include a language within a country different from the native language. In this paper, we do not differentiate between second and foreign languages. Graddol (1997) gives a slightly different definition stating that a second language is an additional language used in some daily contexts more frequently than just a foreign language.

[3]As alternative assumptions on the incentives for second-language acquisition, Ginsburgh et al. (2007) take up the similarity of languages and Caminal (2016) and Caminal and Di Paolo (2019) emphasize the emotional purpose such as acquiring the sense of cultural identity and solidarity.

[4]A "country" throughout this paper is more like a cultural community based on a language than a political nation.

[5]In the three-country case, a different result is obtained than for the case of four or more countries. That result is discussed in Section 5.

(Propositions 5 and 7). If the ratio of gain by the additional communication to effort cost increases, then any equilibrium in which the language of the second, third, or fourth largest country is hegemonic is more likely to exist (Propositions 6 and 7).

Theoretically speaking, this study is in the class of large coalition formation games, with the special feature that an agent can belong to two coalitions through her native and second languages.[6] After obtaining Proposition 2, the model turns out to be a large coordination game in which agents choose a hegemonic language.

Proposition 2 may seem slightly extreme because, in the real world, there are many citizens in many countries who eagerly study languages other than the hegemonic one. This is because of several simplifications in the model such as the payoff structure and that a citizen can learn at most one foreign language. However, this extreme proposition makes us easily focus on the conditions for the stability of a hegemonic language.

Today, one can easily imagine one language as the hegemonic language of the world. That is English. Ethnologue reports in 2019 that by the number of native speakers, English is in the third place with 379 million, following Mandarin with 918 million and Spanish with 460 million.[7] However, by the number of both native and second language speakers, English turns out to be the first with 1,132 million. Concentration on English is observed more clearly in several specific fields. For example, English is universally used for international aviation communications, partly due to the recommendation by the International Civil Aviation Organization (ICAO). Besides, in Japan and some other countries where their native languages are not English, English is primarily used as the tool of domestic aviation communications, even without recommendation by ICAO.[8]

Our model shows the existence of an equilibrium in which a language of a country that does not have the largest population is the hegemonic language. This coincidence with the actual situation in the world suggests that the approach focusing on the population size of speakers is at least worth studying for understanding some aspects of the current linguistic regime in the world. Note that we do not claim the population size of speakers is the sole or the most important reason for the emergence of the hegemonic language. The historical reason why English has become so hegemonic is often considered to be the combination of Britain's colonial expansion and the rise of the United States as the economic and political superpower in the 20th century.[9] We consider the population size as just one factor for

[6]For a survey of recent studies on coalition formation games, see Ray (2008).

[7]See Eberhard et al. (eds., 2019). Mandarin is a dialect or a group of dialects of Chinese.

[8]For a recent survey of economic analysis on English dominance, see Mélitz (2018).

[9]See Graddol (1997). Phillipson (1992) also emphasizes in his influential book that the United Kingdom and the United States maintain the continuous inequalities between

the emergence of hegemonic language and study the mechanism of this factor. We believe that Propositions 5 to 7 briefly explained above might be insightful and worth reporting for the consideration of the actual linguistic situations.

The structure of the rest of the paper is as follows. Section 2 reviews the previous game theoretic models studying the behavior of second-language acquisitions. Section 3 constructs the formal model. Section 4 shows several consequences of analyzing this model. Section 5 discusses several remaining problems. Section 6 concludes. As the proofs for Lemma 1, Remark 1, and Propositions 1 and 2 are rather straightforward, we move them to the online supporting information. As the proofs for Propositions 5 to 8 are technical and long, we move them to the Appendix.

2 Related Literature

Selten and Pool (1991) first construct a large-scaled game model of second-language acquisition behavior. In their model, there are multiple countries and each country has a continuum of agents. An agent has her native language and can learn several foreign languages. She can gain from the additional communication in her foreign languages, but has an effort cost to learn them. Agents within a country can conduct some coalitional actions. Many of such features are taken into our model. While their model is so general that they can state almost only the existence of an equilibrium, our model is more tractable and applicable to the linguistic situations in the world.

Church and King (1993) and Ginsburgh et al. (2007) both construct models with two countries along Selten and Pool (1991)'s approach. Church and King (1993) constructs a simple model in which the equilibrium is slightly extreme, similarly to ours, such that all agents in the smaller country study the foreign language and no one in the larger country studies the other language. Ginsburgh et al. (2007) construct a model for empirical purpose on how the sizes of countries and the similarity of languages affect the behavior of foreign language acquisitions. Our model is different from the others because we construct a more large-scaled picture with four countries and examine properties related to hegemonic language.

Among theoretical studies related to language problems, Caminal (2010) constructs a theoretical model with two major and minor languages and studies how firms produce cultural goods such as books and films in either language. Mélitz (2007) models the competitive market of cultural goods

English and other languages for political and industrial purposes. Graddol (2006) notes that "the English language teaching sector directly earns nearly £1.3 billion for the UK in invisible exports and our other education related exports earn up to £10 billion a year more."

with some empirical evidences and states that almost one-way translation activities from the dominant language to the others take place. The survival of bilingualism or multilingualism within a country has been a topic since the classical Laitin (1993) who theoretically studies the linguistic situation in Ghana employing the coordination game with a tipping point suggested by Schelling (1978). Studies about the survival of bilingualism in a country include Wickström (2005), Iriberri and Uriarte (2012), and Caminal (2016). Clingingsmith (2017) studies how the population of speakers of a language affects its survival, by employing an evolutionary game theory model and empirical studies. He claims that languages with more than 35,000 speakers distribute following a power law in a steady-state, and languages with less than 35,000 speakers are close to extinction.

Recent empirical studies are as follows. Gazzola and Mazzacani (2019) show that English language skills increase the probability of being employed for men in Germany, Italy and Spain and women in Germany and Italy, although French language skills have no significant impact. Asadullah and Xiao (2019) show that English language skills bring a wage premium in China. Foreman-Peck and Zhou (2015) study small and medium size exporting firms in Europe and show the lack of investment in language skills in English native speaking countries. On the other hand, Liwiński (2019) finds that in Poland, skills in Spanish, French or Italian bring a higher wage premium than English skills. Mélitz (2008) studies the role of languages in bilateral trades of the world based on the gravity model. He shows that communication using a common language is important for promoting trade even though direct communication is more effective, and English is not significantly effective among European common languages.

3 The Model

There are four **countries**, X_1, X_2, X_3, and X_4. Let $N := \{1, 2, 3, 4\}$, which is the set of the suffixes of the countries. The suffixes of arbitrary countries are often denoted by a, b, c, and d. In each country, there is a continuum of **agents** (or citizens). **Population** of each country is the length of the continuum of agents. Let x_a denote the population of X_a. We normalize that $x_1 := 1$ and assume that $x_1(= 1) > x_2 > x_3 > x_4$. Each country has a different **language**. Let L_a denote the language of X_a. An agent i in X_a can speak L_a and can learn at most one other language.

Payoff. Active agents are agents who actually study a foreign language. It is endogenously determined whether an agent becomes active in this game. The **net payoff** of an active agent i in X_a learning L_b is $r(x_b + y_b) - t_i$, where r is a constant for all agents in all countries, y_b is the total length of active agents studying L_b who are not in X_a, and t_i is the type of agent i that represents effort cost to study and understand a foreign language. We simply refer to t_i as **cost** for agent i. t_i is independently, identically, and uniformly distributed over $[0, 1]$. We assume that $0 < r < 1/(1+x_2+x_3+x_4)$ to ensure that $0 < r(x_b + y_b) < 1$. Let $u_{ab} := r(x_b + y_b)$, which we call the **gross payoff**, and is the first term of the net payoff. The gross payoff is the same for all active agents in X_a studying L_b. The interpretation of the gross payoff is that the gain from studying a foreign language depends on the length of agents with whom she can communicate in her second language.[10] r represents the ratio of this gain to the cost. If r is large, the gain from communication is large relative to the cost. When i does not study a foreign language, she need not pay her cost t_i, and so her net payoff is normalized to 0.

Timing of action and information revelation. This is a two-stage game. In the first stage, an agent chooses a foreign language. At this stage, an agent does not know the costs for herself and other agents. After the first stage, her cost to study a foreign language is revealed not only to her, but also to all agents. In the second stage, an agent decides whether to study the language chosen in the first stage, *i.e.*, she decides whether to be active or not. All the game structures except the costs are common knowledge among all agents in all countries.

A **coalition** is a continuum of agents in a country with a length smaller than or equal to that population. We assume that in the first stage, if an agent chooses a foreign language, then there is a coalition in the same country that choose the same foreign language. This is because an agent is atomless and her sole choice has no effect in this model.

Equilibrium Concept. We employ a stronger version of the subgame perfect equilibrium as the equilibrium concept, which requires partial stability to some coalitional deviations in the first stage.

In the second stage, all agents' choices should constitute a Nash equilibrium, *i.e.*, any active agent $i \in X_a$ studying L_b should satisfy the following condition:

$$u_{ab} - t_i > 0.$$

We refer to this condition as the **condition in the second stage** (**C2**). Agents who are not active in the second stage do not satisfy **C2**.

In the first stage, based on the prediction for the outcome in the second stage, agents choose languages to maximize their expected net payoffs. Their choices should constitute a Nash equilibrium that is even stable against deviation by any coalition.[11] Remember that when agents form a coalition in the first stage, they do not know each agent's cost, including their own, in

[10] Note that in this model, an agent can increase her payoff by communicating with other agents in a foreign language even when they can speak her native language.

[11] The equilibrium concept in the first stage is a partially "strong Nash equilibrium" presented by Aumann (1959). This partiality means that we only account for a coalition of agents in one country.

the coalition. They only know that agents' costs are uniformly distributed over $[0, 1]$ in the coalition.

There are several possibilities for coalitional deviations in a country, for example, a change in the foreign language study policy by the central government of a country and a boom in some foreign culture and the related foreign language in a country. Thus, it is natural to request stability against coalitional deviations in equilibria.

Hereafter, we refer to this equilibrium concept in this two-stage game as **equilibrium**.

Now, we start preliminary investigations. First, we present how to calculate actual proportions of active agents and their final net payoffs when decisions by all agents in the first stage are given. Let $a, b, c, d \in N$. Let $z_{ab} \in [0, x_a]$ be a coalition in X_a in which members choose to study L_b in the first stage. Let $p_{ab} \in [0, 1]$ be the proportion of active agents in z_{ab}.

C2 requires that the net payoff for an active agent i with t_i should be larger than 0. Let $\bar{t}_{ab} := u_{ab}$, which is the threshold cost for agents in X_a to decide whether to learn L_b in the second stage. If t_i is smaller than $\bar{t}_{ab}$, then an agent i actually studies L_b. Otherwise, she does not study L_b in the second stage. As we assume that t_i is uniformly distributed over $[0, 1]$, the proportion of agents whose costs are smaller than $\bar{t}_{ab}$ is $\bar{t}_{ab}$, *i.e.*, $p_{ab} = \bar{t}_{ab}$. Thus, we have $p_{ab} = u_{ab}$.

Note that the gross payoff of agents in X_a studying L_b is $u_{ab} = r(x_b + p_{cb}z_{cb} + p_{db}z_{db})$. Thus, we also have $u_{ab} = r(x_b + u_{cb}z_{cb} + u_{db}z_{db})$. Similarly, we have $u_{cb} = r(x_b + u_{ab}z_{ab} + u_{db}z_{db})$ and $u_{db} = r(x_b + u_{ab}z_{ab} + u_{cb}z_{cb})$. By solving these three equations simultaneously, the gross payoff of active agents in X_a studying L_b is

$$u_{ab} = \frac{rx_b(rz_{cb} + 1)(rz_{db} + 1)}{1 - r^2(z_{ab}z_{cb} + z_{cb}z_{db} + z_{db}z_{ab}) - 2r^3 z_{ab}z_{cb}z_{db}}.$$

Hence, u_{ab} is a function of z_{ab}, z_{cb} and z_{db}. When the functional form is appropriate, we write $u_{ab}[z_{ab}, z_{cb}, z_{db}]$ to represent the above formula.[12]

The next lemma states some properties of u_{ab} that are useful in the following investigation.

Lemma 1. Let $a, b, c, d \in N$. Let an agent i in X_a choose L_b in the first stage. (i) If z_{cb} increases, then u_{ab} increases. (ii) If z_{ab} increases and $z_{cb} > 0$ or $z_{db} > 0$, then u_{ab} increases.[13]

It is easy to interpret Lemma 1. As we assume that the gross payoff of studying a foreign language depends on the length of agents with whom she

[12]The first variable of u_{ab} should be the proportion of agents in X_a choosing L_b in the first stage. Because in the formula, z_{cb} and z_{db} are symmetric, we need not consider the order of the second and third variables of the functions.

[13]We can define u_{ab} even if $z_{ab} = 0$.

can communicate in her second language, (i) of Lemma 1 is straightforward. The increase of agents in the same country choosing the same foreign language in the first stage may not directly affect her gross payoff. At first, it increases the proportions of active agents studying the same foreign language in other countries. Then, it increases the length of agents with whom she can communicate, and increases her gross payoff, as a consequence, which is (ii) of Lemma 1.

Now, we consider the equivalent condition for an equilibrium in the first stage. The expected net payoff of agent i in X_a deciding to study L_b in the first stage is $\int_0^{\bar{t}_{ab}} \{u_{ab} - t_i\} dt_i$. This is because if $t_i < \bar{t}_{ab}$, i will choose to be active in the second stage, otherwise she does not choose to be active and her payoff is 0. This is also because t_i is uniformly distributed over $[0, 1]$. Then, the expected net payoff can be rewritten as

$$\int_0^{\bar{t}_{ab}} \{u_{ab} - t_i\} dt_i = \int_0^{u_{ab}} \{u_{ab} - t_i\} dt_i = [u_{ab} t_i - \frac{t_i^2}{2}]_0^{u_{ab}} = \frac{u_{ab}^2}{2}.$$

Since $u_{ab} > u'_{ab} \iff u_{ab}^2/2 > u'^2_{ab}/2$, we only need to compare the gross payoffs to confirm an equilibrium. Hence, in the first stage of an equilibrium, for any coalition $0 < z'_{ac} \leq x_a$,

$$u_{ab}[z_{ab}, z_{cb}, z_{db}] \geq u_{ac}[z'_{ac}, z_{bc}, z_{dc}].$$

We refer to this condition as the **condition in the first stage** (**C1**). What we are actually interested in are the choices of agents in the first stage. Hence, we rather focus on the first stage and **C1** in the following investigation.

4 Results

4.1 properties of equilibria

In this section, we derive all equilibria and investigate their properties. First, we define two notions related to equilibria.

A **countrywide coalition** (**CC-**) **equilibrium** is an equilibrium in which all agents in each country choose the same language in the first stage. A **CC-deviation** is a deviation from an equilibrium by all agents in a country. The next remark is rather obvious; however, it is quite useful in the following investigation.

Remark 1. To qualify as an equilibrium, we only need to check **C1** against CC-deviations, *i.e.*, $u_{ab}[z_{ab}, z_{cb}, z_{db}] \geq u_{ac}[x_a, z_{bc}, z_{dc}]$, where z_{ab}, z_{cb}, z_{db}, z_{bc}, and z_{dc} are coalitions in an equilibrium.

The following is the first proposition in this model.

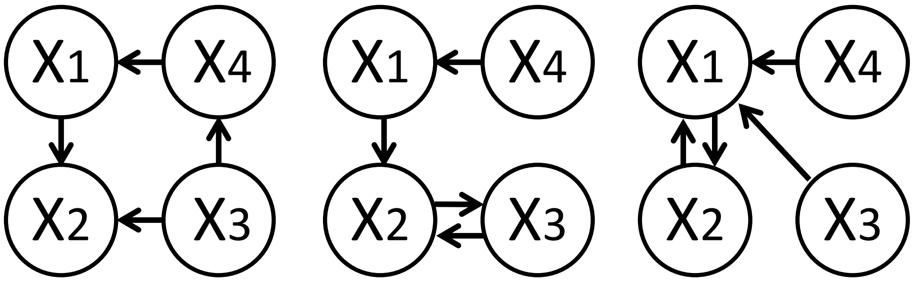

Figure 1: The left side is an example of Situation 2, the center is that of Situation 2, and the right side is that of Situation 3.

Proposition 1. *An equilibrium is always a CC-equilibrium.*

Remember that Lemma 1 (ii) states that agents in a country obtain higher payoffs by forming a larger coalition. This induces Remark 1 and Proposition 1, suggesting that we only need to take into consideration the biggest coalition, which is **CC**. Proposition 1 also suggests that because any coalitional deviation is possible, a situation where agents in a country study different languages is quite unstable in this model.

Now, we can summarize situations in terms of directed graph theory.[14] Suppose that countries are **vertices** and the choices of agents of those countries in the first stage are represented by **arcs**. A home country of agents is represented by the **tail** of the arc, and the country whose language is chosen by the agents is represented by the **head** of the arc.

By Proposition 1, there are four arcs and each country must be the tail of only one arc. The number of possible heads of each arc is three. Thus, there is the possibility of $3^4 = 81$ situations (or graphs).

We divide the situations into three types. **Situation 1** is that any country is head of at most one arc; **Situation 2** is that any country is head of at most two arcs; and **Situation 3** is that there is a country that is head of three arcs. See Figure 1 for the illustrations. Then, we have the following proposition.

Proposition 2. *Situations 1 and 2 do not occur in an equilibrium.*

In both situations, without loss of generality, while all agents in X_a choose L_b, there is X_c that agents do not choose L_b. Then, we can show

[14]We borrow some elementary terms of directed graph theory to represent situations and do not demand knowledge of directed graph theory. If recent developments in directed graph theory are of interest, see Bang-Jensen and Gutin (2000) for a survey. The terms of graph theory used in this paper follow them. For applications of graph theory to social sciences, see Jackson (2010).

that agents in X_a or those in X_c have incentive to CC-deviation, which is a contradiction to **C1**.

By Propositions 1 and 2, if an equilibrium exists, then it belongs to Situation 3. Actually, there are four equilibria in this model and all belong to Situation 3 as long as certain conditions are satisfied.

A language is called **hegemonic** if all agents in all other countries choose it in the first stage. Because any case of Situation 3 has a hegemonic language, there always exists an equilibrium. We can show that an equilibrium in which each of the four languages is hegemonic is unique. We refer to the equilibrium in which L_a is hegemonic as **hegemony** a (**H_a-**) **equilibrium**.

The next proposition states that, in this model, there is an equilibrium without any condition.

Proposition 3. *There is H_1-equilibrium in which all agents in X_2, X_3, and X_4 choose L_1, and all agents in X_1 choose L_2 in the first stage.*

Proof of Proposition 3. We show that H_1-equilibrium surely satisfies **C1**. An active agent in X_1 has gross payoff $u_{12}[x_1, 0, 0] = rx_2$ in the equilibrium. Consider $a \in N\backslash\{1, 2\}$. Suppose that agents in X_1 CC-deviate to X_a. Then, an active agent X_1 has gross payoff $u_{1a}[x_1, 0, 0] = rx_a$. Because $x_2 > x_a$, **C1** for agents in X_1 is satisfied.

Let $b, c, d \in N\backslash\{1\}$. In equilibrium, an active agent in X_b has net payoff $u_{b1}[x_b, x_c, x_d] = \{r(rx_c+1)(rx_d+1)\}/\{1-r^2(x_bx_c+x_cx_d+x_dx_b)-r^3x_bx_cx_d\}$. Suppose that agents in X_b CC-deviate to X_c. Then, an active agent has net payoff $u_{bc}[x_b, 0, 0] = rx_c$. Obviously $u_{b1} > 1$ and $1 > rx_c$, **C1** for agents in X_b is satisfied. □

The existence of H_1-equilibrium without any condition in this model is not surprising. Because agents gain from having more agents to communicate with, it is quite natural that the language in the largest country can be hegemonic and all agents in other countries study this language. However, the hegemonic language has not always been such a language. In the next subsection, we show that even in this basic model, there is an equilibrium in which a language in a small country becomes hegemonic under some conditions.

4.2 Conditions when the hegemonic country is not the largest country

In this subsection, we first focus on the H_2-equilibrium as it is the most possible and tractable equilibrium except the H_1-equilibrium. Later, we show that the H_3- and the H_4-equilibria actually have essentially the same existence conditions.

We first check the conditions under which the H_2-equilibrium exists, and then investigate how changes in the parameters affect the existence of this equilibrium.

Proposition 4. *There is H_2-equilibrium in which all agents in X_1, X_3, and X_4 choose L_2 and all agents in X_2 choose L_1 in the first stage if the gross payoff of agents in X_3 in the equilibrium is larger than or equal to their gross payoff when CC-deviating to study L_1, i.e., if $u_{32}[x_3, x_1(=1), x_4] - u_{31}[x_3, x_2, 0] \geq 0$.*

Proof of Proposition 4. It is obvious that in this equilibrium, agents in X_1 and those in X_2 have no incentive for a CC-deviation. It is also obvious that agents in X_3 have no incentive to CC-deviate to study L_4 and those in X_4 have no incentive to deviate to study L_3. The net payoff of an active agent i in X_3 in the equilibrium is

$$u_{32}[x_3, 1, x_4] = \frac{rx_2(r+1)(rx_4+1)}{1 - r^2(x_3 + x_4 + x_3x_4) - 2r^3x_3x_4}. \tag{1}$$

Her net payoff in a CC-deviation to study L_1 is

$$u_{31}[x_3, x_2, 0] = \frac{r(rx_2+1)}{1 - r^2x_2x_3}. \tag{2}$$

The net payoff of an active agent j in X_4 in the equilibrium is

$$u_{42}[x_4, 1, x_3] = \frac{rx_2(r+1)(rx_3+1)}{1 - r^2(x_3 + x_4 + x_3x_4) - 2r^3x_3x_4}. \tag{3}$$

Her net payoff in a CC-deviation to study L_1 is

$$u_{41}[x_4, x_2, 0] = \frac{r(rx_2+1)}{1 - r^2x_2x_4}. \tag{4}$$

The condition for agents in X_3 staying in the equilibrium is (1) − (2) ≥ 0 and that for agents in X_4 is (3) − (4) ≥ 0. Note that the numerator of (1) is smaller than that of (3) and the denominators of (1) and (3) are equal. Thus, (1) is smaller than (3). Also note that the numerators of (2) and (4) are equal and the denominator of (2) is smaller than that of (4). Thus, (2) is larger than (4). Hence, if (1) − (2) ≥ 0 holds, (3) − (4) ≥ 0 always holds. Hence, we only need (1) − (2) ≥ 0 as the sufficient condition for the H_2-equilibrium. □

We investigate how each parameter of this model affects the existence of the H_2-equilibrium. Of course, in this investigation, the sufficient condition $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0] \geq 0$ plays an important role. Table 1 contains numerical examples of how changes in parameters such as r, x_2, x_3 and x_4 affect the existence of equilibria including the H_2-equilibrium. As Table 1 suggests, we have the following proposition.

Proposition 5. *The H_2-equilibrium is more likely to exist if x_2, x_3, or x_4 increases, in the sense that $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ increases in x_2, x_3, and x_4.*

TABLE 1: Numerical examples of how changes in parameters such as r, x_2, x_3 and x_4 affect the existence of equilibria

r	x_2	x_3	x_4	$u_{32}[x_3, 1, x_4]$	$u_{31}[x_3, x_2, 0]$	$u_{32} - u_{31}$	H_2	H_3	H_4
0.1	0.9	0.8	0.7	0.1083	0.1098	−0.0015			
0.2	0.8	0.7	0.6	0.2326	0.2373	−0.0047			
0.2	0.9	0.8	0.3	0.2429	0.2430	−0.0001			
0.2	0.9	0.8	0.4	0.2497	0.2430	0.0067	E		
0.2	0.9	0.8	0.7	0.2710	0.2430	0.0280	E	E	
0.3	0.8	0.6	0.3	0.3808	0.3888	−0.0080			
0.3	0.8	0.6	0.4	0.3992	0.3888	0.0104	E		
0.3	0.8	0.7	0.6	0.4476	0.3917	0.0559	E	E	
0.3	0.9	0.8	0.6	0.5146	0.4074	0.1072	E	E	E

E: Existence

We can show that the derivative of $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ with respect to x_2 is positive. We can obtain the proofs more straightforwardly in the cases of x_3 and x_4.

It is intuitive that if the population of X_2 increases, the H_2-equilibrium is more likely to exist. Actually, the increase of x_2 positively affects both $u_{32}[x_3, x_1, x_4]$ and $u_{31}[x_3, x_2, 0]$, whose precise expressions are in (1) and (2). Proposition 5 suggests that surely the effect of the increase of x_2 on $u_{32}[x_3, x_1, x_4]$ is larger than that on $u_{31}[x_3, x_2, 0]$.

The fact that the increase of x_3 or x_4 makes the existence of the H_2-equilibrium more likely is rather interesting. x_3 obviously positively affects both $u_{32}[x_3, x_1, x_4]$ and $u_{31}[x_3, x_2, 0]$ since the CC-deviation by the agents in X_c itself is the key for the condition of the existence of the equilibrium. It is shown that the effect of x_3 on $u_{32}[x_3, x_1, x_4]$ is larger. x_4 positively affects only $u_{32}[x_3, x_1, x_4]$. Overall, increases in the populations of relatively small countries whose agents choose the hegemonic language make the existence of the H_2-equilibrium more likely.

The following proposition states that essentially, if r increases, then the H_2-equilibrium is more likely to exist.

Proposition 6. *The H_2-equilibrium is more likely to exist if r increases, in the sense that, given r, r' such that $r < r'$, if the H_2-equilibrium exists in r, then it exists in r'.*

It is sufficient to show that the second derivative of $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ with respect to r is positive. Figure 2 depicts typical graphs of $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ with respect to r. The outcome of this function is smaller than 0 when r is close to 0 as the first derivative of $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ with respect to r is negative when r is close to 0.

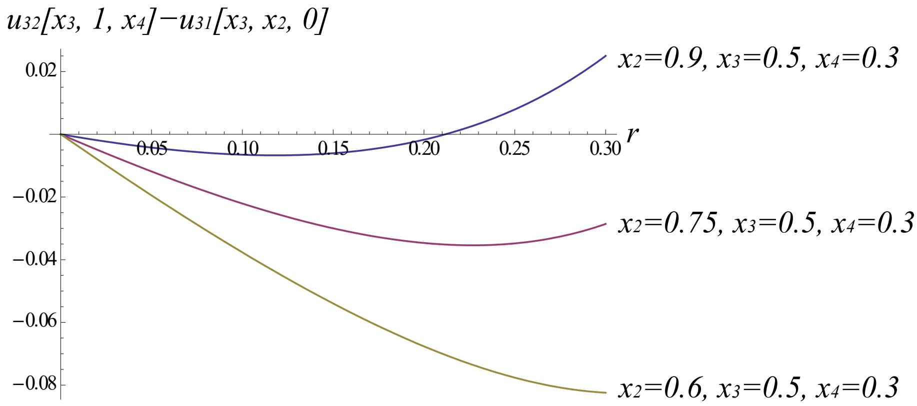

FIGURE 2: Typical graphs of $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ with respect to r

r is interpreted as the ratio of gain by the additional communication by the learned hegemonic language relative to effort cost. Thus, if r becomes larger, the number of actual agents in each country increases. This makes the existence of the H_2-equilibrium more likely. This proposition may be interesting with respect to language policy. If the country with the hegemonic language wishes to maintain its position, a policy of reducing the language learning cost is effective.

Now, we focus on the H_3- and the H_4-equilibria. The next proposition is rather surprising. The essentially symmetric arguments to the existence of the H_2-equilibrium hold on both the H_3- and the H_4-equilibria.

Proposition 7. *Let $a, b \in \{3, 4\}$. There is H_a-equilibrium in which all agents in X_1, X_2, and X_b choose L_a and all agents in X_a choose L_1 in the first stage if the gross payoff of agents in X_2 in the equilibrium is larger than or equal to their gross payoff when deviating to study L_1, i.e., if $u_{2a}[x_2, x_1(= 1), x_b] - u_{21}[x_2, x_a, 0] \geq 0$.*

The sufficient condition $u_{2a}[x_2, 1, x_b] - u_{21}[x_2, x_a, 0] \geq 0$ and the symmetric argument to Proposition 5 suggest that the increase of x_2 brings the emergence of the H_a-equilibrium. However, if x_2 increases, $u_{1a}[1, x_2, x_b] - u_{12}[1, 0, 0]$ decreases and the CC-deviation by agents in X_1 to study L_2 may possibly occur. Proposition 7 states that we need not consider the latter condition. If the sufficient condition of this proposition $u_{2a}[x_2, x_1, x_b] - u_{21}[x_2, x_a, 0] \geq 0$ holds, then $u_{1a}[1, x_2, x_b] - u_{12}[1, 0, 0] \geq 0$ holds.

Note that in the proofs for Propositions 6 and 7, we never use the fact that $x_2 > x_3$ and $x_2 > x_4$. Hence, the symmetric statements to Propositions 5 and 6 also hold on both the H_3- and the H_4-equilibria. Increases of x_2, x_3, x_4, and r contribute to the existence of the H_3- and the H_4-equilibria.

5 Discussion

5.1 The effect of the number of countries

In this article, we have investigated the model of second-language choice in four countries. In this subsection, we discuss what happens when the number of countries changes.

First, we note the three-country case. In this case, there is only one equilibrium, which is the H_1-equilibrium.

Proposition 8. *Suppose that there are three countries X_1, X_2, and X_3. Then, there is a unique equilibrium. It is the H_1-equilibrium in which all agents in X_2 and X_3 choose L_1, and all agents in X_1 choose L_2 in the first stage.*

The intuition behind the proof is as follows. Even in the three country case, Proposition 1 holds and an equilibrium should be a CC-equilibrium by the similar argument from the four country case. As an example, consider the situation in which all agents in X_1 and X_3 study L_2 and all agents in X_2 study L_1 in the first stage. This corresponds with the H_2-equilibrium in the four country case. This is not an equilibrium because agents in X_3 have incentive to CC-deviate to study L_1 and constitute the equilibrium of this case. The key factor of the incentive for CC-deviation by agents in X_3 is that the proportion of active agents in X_1 studying L_2 in this situation is smaller than the proportion of active agents in X_2 studying L_1 in the equilibrium. In the H_2-equilibrium of the four country case, if x_4 is sufficiently large, then the size of L_2 speakers also becomes large and the condition for the existence of the equilibrium in Proposition 4 is satisfied.

On the other hand, if the number of countries is greater than or equal to five, the investigation and all propositions for the case of four countries hold.

5.2 Welfare analysis

As has already been mentioned when defining net payoff, the actual utility of an agent i in X_a depends not only on her net payoff but also on y_a, the number of foreign agents with whom she can communicate in her own language. If we take this into account, it is obvious that all four equilibria in this model are Pareto efficient because an agent with quite a high cost never learns a foreign language, and she can only increase her actual utility if her native language becomes hegemonic.

5.3 An extension to mixed strategies

Here, consider a simple extension of this model to allow mixed strategies in the first stage. We assume that in the first stage, if an agent chooses a mixed strategy over the set of foreign languages, then there is a coalition that choose the same mixed strategy. In this extension, there is no mixed-strategy equilibrium in the first stage. The reason is as follows. If a mixed-strategy equilibrium exists, then there are a positive proportion of agents who take the same expected net payoffs from at least two languages. Then a coalitional deviation of them to give probability 1 on choosing one of these languages brings them higher payoffs, which is a contradiction. In this extension, all propositions in this paper hold.

5.4 The ratio of gain to cost reconsidered

We have assumed throughout the paper that r represents the ratio of the gain from understanding a foreign language and communicating with other agents in this additional language to its effort cost. In the real world, r varies among linguistic relations. For example, native English speakers can learn French or Spanish more easily than Chinese, Korean, or Japanese. Moreover, because possessing the hegemonic language is usually considered to bring welfare improvement for agents in the country, it may often promote campaigns to increase the gain/cost ratio to learn their language for foreigners via media and schools. The targets of these campaigns are sometimes not all countries, but one country in particular.

Suppose that for an agent in X_1, the learning cost of L_4 is lower than those of other foreign languages, and vice versa. For an agent in X_2, the learning cost of L_3 is lower than those of other foreign languages, and vice versa. Then, our model with this extension easily suggests the emergence of two blocks of languages and the disappearance of a hegemonic language. This is an interesting direction of analysis, especially when considering the language policies of a country.

6 Concluding remarks

In this paper, we have developed a large-scale game model of second-language acquisition. In an equilibrium for this model, a hegemonic language always exists and we have studied the conditions for the existence of a hegemonic language of which the number of native speakers is not the largest. As we have employed several simplified assumptions, such as the fact that the gain of an agent is determined only by the number of agents with whom she can communicate in the additionally learned language, there remain many important features of second-language acquisition that have not been studied. In the present model, agents in a country learn the same language, even though in the real world, a significant number of people in various countries study several minor languages for several purposes. This model is quite static, and lacks the dynamics of convergence to an equilibrium with one hegemonic language and its collapse. However, we believe that this tractable model contains several interesting implications and will stimulate further research on sociolinguistics via formal modelling approaches.

Appendix

Proof of Proposition 5. $u_{32}[x_3, x_1(=1), x_4] - u_{31}[x_3, x_2, 0]$ is explicitly written as

$$\begin{aligned}
&u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0] \\
=&\{r(x_2 - 1) + r^2 x_2 x_4 + r^3(-x_2^2 x_3 + x_2 x_4 + x_3 + x_3 x_4 + x_4) \\
&+ r^4(x_2 x_3 - x_2^2 x_3 + x_2 x_3 x_4 - x_2^2 x_3 x_4 + x_2 x_4 + 2x_3 x_4) \\
&+ r^5(2x_2 x_3 x_4 - x_2^2 x_3 x_4)\} \\
&/\{(1 - r^2 x_2 x_3)(1 - r^2 x_3 - r^2 x_4 - r^2 x_3 x_4 - 2r^3 x_3 x_4)\}. \qquad (5)
\end{aligned}$$

1. By differentiating (5) with respect to x_2, we have that

$$\begin{aligned}
&\{r + r^2 x_4 + r^3(-2x_2 x_3 - x_3 + x_4) \\
&+ r^4(-2x_2 x_3 - 2x_2 x_4 x_3 + x_4 x_3 + x_3 + x_4) \\
&+ r^5(x_2^2 x_3^2 + x_4 x_3^2 + x_3^2 - 2x_2 x_3 x_4 + 3x_3 x_4) \\
&+ r^6(x_2^2 x_3^2 + x_2^2 x_4 x_3^2 + 2x_4 x_3^2) + r^7 x_2^2 x_3^2 x_4\} \\
&/\{(1 - r^2 x_2 x_3)^2 (1 - r^2 x_3 - r^2 x_3 x_4 - r^2 x_4 - 2r^3 x_3 x_4)\}. \qquad (6)
\end{aligned}$$

The denominator of (6) is obviously positive. By the definition of r, $r < 1/(1+x_2+x_3+x_4) \iff 1/r > (1+x_2+x_3+x_4)$. Thus, $r > r^3(1+x_2+x_3+x_4)^2 > r^3(1+x_2+x_3+x_4)+r^4(1+x_2+x_3+x_4)^2(x_2+x_3+x_4)$. As $x_2, x_3, x_4 < 1$, the right-hand side of this inequality is larger than $-\{r^3(-2x_2x_3 - x_3) + r^4(-2x_2x_3 - 2x_2x_4)\}$. Also note that $r^5(-2x_2x_3x_4 + 3x_3x_4) > 0$ since $x_2 < 1$. Thus, the numerator of (6) is positive. Hence, (6) is positive and (5) is increasing in x_2.
2. Suppose that x_3 increases. The two factors of the denominator of (5) $(1 - r^2x_2x_3)$ and $(1 - r^2x_3 - r^2x_4 - r^2x_3x_4 - 2r^3x_3x_4)$ both decrease. Thus, the denominator of (5) decreases. Since $x_2, x_4 < 1$, $\{-x_2^2x_3 + x_3\}$, $\{x_2x_3 + -x_2^2x_3 + x_2x_3x_4 - x_2^2x_3x_4\}$, and $\{x_2x_3x_4 - x_2^2x_3x_4\}$ increase. Thus, the numerator of (5) increases. Hence, (5) increases.
3. Suppose that x_4 increases. Then, $\{1 - 2r^3x_3x_4 - r^2x_3 - r^2x_4 - r^2x_3x_4\}$ decreases. Thus, the denominator of (5) decreases. Now $x_2, x_3 < 1$, $\{x_2x_3x_4 - x_2^2x_3x_4\}$ and $\{x_2x_3x_4 - x_2^2x_3x_4\}$ increase. Thus, the numerator of (5) increases. Hence, (5) increases. □

Proof of Proposition 6. We show that the second derivative of $\{u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]\}$ with respect to r is positive. Let F_N denote the numerator of (1), F_D denote the denominator of (1), S_N denote the numerator of (2), and

S_D denote the denominator of (2). Then, $\{u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]\} = \{F_N/F_D - S_N/S_D\}$, and its second derivative is

$$
\begin{aligned}
&\{\frac{2F_N F_D'^2}{F_D^3} - \frac{2F_N' F_D'}{F_D^2} - \frac{F_N F_D''}{F_D^2} + \frac{F_N''}{F_D}\} \\
&-\{\frac{2S_N S_D'^2}{S_D^3} - \frac{2S_N' S_D'}{S_D^2} - \frac{S_N S_D''}{S_D^2} + \frac{S_N''}{S_D}\} \\
=&\{\frac{2F_N' F_D'^2}{F_D^3} - \frac{2S_N S_D'^2}{S_D^3}\} + \{-\frac{2F_N' F_D'}{F_D^2} + \frac{2S_N' S_D'}{S_D^2}\} \\
&+\{-\frac{F_N F_D''}{F_D^2} + \frac{S_N S_D''}{S_D^2}\} + \{\frac{F_N''}{F_D} - \frac{S_N''}{S_D}\}. \qquad (7)
\end{aligned}
$$

We show that all four terms in the parentheses of (7) are positive. First, note that $F_N > 0$, $F_D > 0$, $S_N > 0$, $S_D > 0$, and, particularly, that $F_D < S_D$. Also note that $F_N' = x_2 + 2rx_2 + 2rx_2x_4 + 3r^2x_2x_4 > 0$, $F_N'' = 2x_2 + 2x_2x_4 + 6rx_2x_4 > 0$, $F_D' = -2rx_3 - 2rx_4 - 2rx_3x_4 - 6r^2x_3x_4 < 0$, $F_D'' = -2x_3 - 2x_4 - 2x_3x_4 - 12rx_3x_4 < 0$, $S_N' = 1 + 2rx_2 > 0$, $S_N'' = 2x_2 > 0$, $S_D' = -2rx_2x_3 < 0$, and $S_D'' = -2x_2x_3 < 0$. Thus, it is sufficient to show that (i) $F_N' F_D'^2 > S_N S_D'^2$, (ii) $F_N' F_D' < S_N' S_D'$, (iii) $F_N F_D'' < S_N S_D''$, and (iv) $F_N'' > S_N''$. Note that in (ii) and (iii), both sides of the inequalities are negative.
(i) $F_N F_D'^2 - S_N S_D'^2 = \{rx_2(r+1)(rx_4+1)\}\{-2rx_3 - 2rx_4 - 2rx_3x_4 - 6r^2x_3x_4\}^2 - \{r(rx_2+1)\}\{-2rx_2x_3\}^2 > rx_2(r+1)(-2rx_3)^2 - \{r(rx_2+1)\}\{-2rx_2x_3\}^2 = 4r^3(r+1)x_2x_3^2 - 4r^3(rx_2+1)x_2^2x_3^2 > 0$.
(ii) $F_N' F_D' - S_N' S_D' = (x_2 + 2rx_2 + 2rx_2x_4 + 3r^2x_2x_4)(-2rx_3 - 2rx_4 - 2rx_3x_4 - 6r^2x_3x_4) - (1 + 2rx_2)(-2rx_2x_3) < (x_2 + 2rx_2)(-2rx_3) - (1 + 2rx_2)(2rx_2x_3) = -4rx_2(1 + r + rx_2)x_3 < 0$.
(iii) $F_N F_D'' - S_N S_D'' = \{rx_2(r+1)(rx_4+1)\}\{-2x_3 - 2x_4 - 2x_3x_4 - 12rx_3x_4\} - \{r(rx_2+1)\}\{-2x_2x_3\} < rx_2(r+1)(-2x_3) - r(rx_2+1)(-2x_2x_3) = 2r^2(-1 + x_2)x_2x_3 < 0$.
(iv) $F_N'' - S_N'' = (2x_2 + 2x_2x_4 + 6rx_2x_4) - (2x_2) > 0$. □

Proof of Proposition 7. First, we consider the case of $a = 3$ and $b = 4$. We only show that if $u_{23}[x_2, 1, x_4] - u_{21}[x_2, x_3, 0] \geq 0$ holds, then $u_{13}[1, x_2, x_4] - u_{12}[1, 0, 0] \geq 0$ holds. By the similar argument to the proof of Proposition 4, we can easily find that the sufficient condition of this proposition implies any other **C1**.

$$
\begin{aligned}
&u_{23}[x_2, 1, x_4] - u_{21}[x_2, x_3, 0] \\
=&\frac{rx_3(r+1)(rx_4+1)}{1 - r^2(x_2 + x_4 + x_2x_4) - 2r^3x_2x_4} - \frac{r(rx_3+1)}{1 - r^2x_2x_3} \\
=&\{rx_3(r+1)(rx_4+1)(1 - r^2x_2x_3) \\
&- r(rx_3+1)(1 - r^2(x_2 + x_4 + x_2x_4) - 2r^3x_2x_4)\}
\end{aligned}
$$

$$/\{(1-r^2(x_2+x_4+x_2x_4)-2r^3x_2x_4)(1-r^2x_2x_3)\} \tag{8}$$

We also have that

$$\begin{aligned}
&u_{13}[1,x_2,x_4]-u_{12}[1,0,0]\\
=&\frac{rx_3(rx_2+1)(rx_4+1)}{1-r^2(x_2+x_4+x_2x_4)-2r^3x_2x_4}-rx_2\\
=&\{rx_3(rx_2+1)(rx_4+1)-rx_2(1-r^2(x_2+x_4+x_2x_4)-2r^3x_2x_4)\}\\
&/\{1-r^2(x_2+x_4+x_2x_4)-2r^3x_2x_4\}.
\end{aligned} \tag{9}$$

We assume that (8) is larger than or equal to 0. As the denominator of (8) is obviously larger than 0, the numerator of (8) is larger than or equal to 0. By using this inequality, we can show that the numerator of (9) is greater than or equal to 0. In this calculation, we need to refer to the definition of r. Then, we obtain that (9) is larger than or equal to 0.

As we do not use the fact that $x_3 > x_4$ to prove the case of $a = 3$ and $b = 4$, the same argument can be applied to the case of $a = 4$ and $b = 3$. □

Proof of Proposition 8. Let $a, b, c \in N$. By solving two gross payoffs $u_{ab} = r(x_b + u_{cb}z_{cb})$ and $u_{cb} = r(x_b + u_{ab}z_{ab})$ simultaneously, we have the gross payoff function of z_{ab} and z_{cb} as

$$u_{ab}[z_{ab}, z_{cb}] = \frac{rx_b(rz_{cb}+1)}{1-r^2z_{ab}z_{cb}}.$$

By similar arguments to those used in the case with four countries, we can obtain Lemmas 1 and Proposition 1. Thus, we only take into account CC-equilibria. For $a \in \{2,3\}$, $b \in \{1,3\}$, and $c \in \{1,2\}$, let $S(a,b,c)$ denote the situation in which all agents in X_1 choose to study L_a, all agents in X_2 choose to study L_b, and all agents in X_3 choose to study L_c. As S can represent all prospective CC-equilibria, the number of possibilities is $2^3 = 8$. $S(2,1,1)$, which is the H_1-equilibrium in the three-countries case, is obviously an equilibrium. We show that the other seven situations are not equilibria.

In the following three situations, agents in X_3 have incentive to CC-deviate to study L_1. In $S(3,1,2)$, we have $u_{32}[x_3,0] - u_{31}[x_3,x_2] = rx_2 - \{r(rx_2+1)\}/\{1-r^2x_2x_3\} < 0$. In $S(3,3,2)$, we have $u_{32}[x_3,0]-u_{31}[x_3,0] = rx_2 - r < 0$. In $S(2,1,2)$, we have $u_{32}[x_3,1]-u_{31}[x_3,x_2] = \{rx_2(r+1)\}/\{1-r^2x_3\} - \{r(rx_2+1)\}/\{1-r^2x_2x_3\}$

$$= \frac{r(x_2-1)(1-r^2x_2x_3-r^2x_3-r^3x_2x_3)}{(1-r^2x_2x_3)(1-r^2x_3)} < 0.$$

In the following three situations, agents in X_2 have incentive to CC-deviate to study L_1. In $S(2,3,1)$, we have $u_{23}[x_2,0] - u_{21}[x_2,x_3] = rx_3 -$

$\{r(rx_3+1)\}/\{1-r^2x_2x_3\} < 0$. In $S(2,3,2)$, we have $u_{23}[x_2,0]-u_{21}[x_2,0] = rx_3-r < 0$. In $S(3,3,1)$, we have $u_{23}[x_2,1]-u_{21}[x_2,x_3] = \{rx_3(r+1)\}/\{1-r^2x_2\} - \{r(rx_3+1)\}/\{1-r^2x_2x_3\}$

$$= \frac{r(x_3-1)(1-r^2x_2x_3-r^2x_3-r^3x_2x_3)}{(1-r^2x_2x_3)(1-r^2x_2)} < 0.$$

In $S(3,1,1)$, agents in X_1 have incentive to CC-deviate to study L_2 since $u_{13}[1,0]-u_{12}[1,0] = rx_3 - rx_2 < 0$. □

Acknowledgement. The author thanks Atsushi Kajii, Keisuke Kawata, Yasuhiro Sato, Takashi Ui, an associate editor and especially an anonymous reviewer for helpful suggestions and comments. Any errors are the author's responsibility. The author declares no conflicts of interest associated with this study. This study received no specific grant from any funding agency. The title of this study was formerly "A Game Theoretic Analysis of Second-language Acquisition and the World's Language Regime".

References

Anderson, Benedict (1983), "Imagined Communities: Reflections on the Origin and Spread of Nationalism", Verso, London.

Asadullah, M. Niaz; Xiao, Saizi, Labor Market Returns to Education and English Language Skills in the People ' s Republic of China: An Update, Asian Development Review 36 (1), 80-111. DOI: 10.1162/adev_a_00124

Aumann, Robert (1959), Acceptable Points in General Cooperative N-Person Games, in A. W. Tucker, R. D. Luce (eds.) "Contributions to the Theory of Games IV", Princeton University Press, Princeton.

Bang-Jensen, Jorgen; Gutin, Gregory (2000), "Digraphs: Theory, Algorithms and Applications", Springer-Verlag, Berlin.

Bleakley, Hoyt; Chin, Aimee (2004), Language Skills and Earnings: Evidence from Childhood Immigrants, Review of Economics and Statistics 86 (2), 481-496. DOI: 10.1162/003465304323031067

Budría, Santiago; Colino, Alberto; de Ibarreta, Carlos M. (2019) The Impact of Host Language Proficiency on Employment Outcomes among Immigrants in Spain, Empirica (2019) 46 (4), 625-652. DOI: 10.1007/s10663-018-9414-x

Caminal, Ramon (2010), Markets and Linguistic Diversity, Journal of Economic Behavior and Organization 76 (3), 774-790. DOI: 10.1016/j.jebo.2010.09.009

Caminal, Ramon (2016), The Economic Value of Reciprocal Bilingualism, in M. Gazzola, B. A. Wickström (eds.) "The Economics of Language Policy", MIT Press, Cambridge.

Caminal, Ramon; Di Paolo, Antonio (2019), Your Language or Mine? The Non-Communicative Benefits of Language Skills, Economic Inquiry 57 (1), 726-750. DOI: 10.1111/ecin.12542

Chiswick, Barry R.; Miller, Paul W. (2010), Occupational Language Requirements and the Value of English in the US Labor Market, Journal of Population Economics 23 (1), 353372. DOI 10.1007/s00148-008-0230-7

Church, Jeffrey; King, Ian (1993), Bilingualism and Network Externalities, Canadian Journal of Economics 26 (2), 337-345. URL: http://www.jstor.org/stable/135911

Clingingsmith, David (2017), Are the World's Languages Consolidating? The Dynamics and Distribution of Language Populations, Economic Journal 127 (599), 143-176. DOI: 10.1111/ecoj.12257

Dustmann, Christian; Fabbri, Francesca (2003), Language Proficiency and Labour Market Performance of Immigrants in the UK, Economic Journal 113 (489), 695-717. DOI: 10.1111/1468-0297.t01-1-0015

Eberhard, David M.; Simons, Gary F.; Fennig, Charles D. (eds.) (2019) "Ethnologue: Languages of the World, 22nd Edition", SIL International, Dallas. URL: http://www.ethnologue.com

Foreman-Peck, James; Zhou, Peng (2015), Firm-Level Evidence for the Language Investment Effect on SME Exporters, Scottish Journal of Political Economy 62 (4), 351-377. DOI: 10.1111/sjpe.12072

Gazzola, Michele; Mazzacani, Daniele (2019), Foreign Language Skills and Employment Status of European Natives: Evidence from Germany, Italy and Spain, Empirica 46 (4), 713-740. DOI: 10.1007/s10663-019-09460-7

Ginsburgh, Victor; Ortuño-Ortín, Ignacio; Weber, Shlomo (2007), Learning Foreign Languages: Theoretical and Empirical Implications of the Selten and Pool model, Journal of Economic Behavior and Organization 64 (3-4), 337-347. DOI: 10.1016/j.jebo.2006.10.005

Graddol, David (1997), "The Future of English?: A Guide to Forecasting the Popularity of the English Language in the 21st Century", British Council, London.

Graddol, David (2006), "English Next: Why Global English May Mean the End of English as a Foreign Language", British Council, London.

Iriberri, Nagore; Uriarte, José-Ramón (2012), Minority Language and the Stability of Bilingual Equilibria, Rationality and Society 24 (4), 442-462. DOI: 10.1177/1043463112453556

Jackson, Matthew O. (2010), "Social and Economic Networks", Princeton University Press, Princeton.

Laitin, David D. (1993), The Game Theory of Language Regimes, International Political Science Review 14 (3), 227-239. DOI: 10.1177/019251219301400302

Liwiński, Jacek (2019), The Wage Premium from Foreign Language Skills, Empirica 46 (4), 691-711. DOI: 10.1007/s10663-019-09459-0

Mélitz, Jacques (2007), The Impact of English Dominance on Literature and Welfare, Journal of Economic Behavior and Organization 64 (2), 193-215. DOI: 10.1016/j.jebo.2006.10.003

Mélitz, Jacques (2008), Language and Foreign Trade, European Economic Review 52 (4), 667-699. DOI: 10.1016/j.euroecorev.2007.05.002

Mélitz, Jacques (2018), English as a Lingua Franca: Facts, Benefits and Costs, World Economy 41 (7), 1747-1916. DOI: 10.1111/twec.12643

Phillipson, Robert (1992), "Linguistic Imperialism", Oxford University Press, Oxford.

Ray, Debraj (2008), "A Game-Theoretic Perspective on Coalition Formation (The Lipsey Lectures)", Oxford University Press, Oxford.

Schelling, Thomas C. (1978), "Micromotives and Macrobehavior", Norton, New York.

Selten, Reinhard; Pool, Jonathan (1991), The Distribution of Foreign Language Skills as a Game Equilibrium, in R. Selten (ed.), "Game Equilibrium Models IV: Social and Political Interaction", Springer-Verlag, Berlin.

Tainer, Evelina (1988), English Language Proficiency and the Determination of Earnings among Foreign-Born Men, Journal of Human Resources 23 (1), 108-122. URL: http://www.jstor.org/stable/145847

Wickström, Bengt-Arne (2005), Can Bilingualism be Dynamically Stable?: A Simple Model of Language Choice, Rationality and Society 17 (1), 81-115. DOI: 10.1177/1043463105051776

Supporting Information for "Second-Language Acquisition Behavior and Hegemonic Language"

Kentaro HATSUMI
Faculty of Policy Studies, Kansai University, Suita, Osaka, 564-8680, Japan
October 5, 2021

This note contains the precise proofs for Lemma 1, Remark 1 and Propositions 1 and 2 that are omitted in the main article.

Proof of Lemma 1. (i) Suppose that z_{cb} increases. Then, the numerator of p_{ab} increases since $r > 0$, $x_b > 0$ and $z_{db} \geq 0$. If $z_{ab} > 0$ or $z_{db} > 0$, the denominator of p_{ab} decreases, otherwise it is unchanged. Thus, p_{ab} increases.

(ii) Suppose that z_{ab} increases. The numerator of p_{ab} is unchanged. If $z_{cb} > 0$ or $z_{db} > 0$, the denominator of p_{ab} decreases, otherwise it is unchanged. Thus, p_{ab} increases as long as $z_{cb} > 0$ or $z_{db} > 0$. □

Proof of Remark 1. By Lemma 1 (ii), a CC-deviation brings a higher or equal payoff to a deviation with a smaller coalition for an agent in it. □

Proof of Proposition 1. Suppose that there is an equilibrium that is not a CC-equilibrium. Then a positive proportion z_{ab} of agents in X_a chooses L_b and another positive proportion z_{ac} chooses L_c. **C1** implies that $u_{ab} = u_{ac}$.

Note that there is a positive proportion of active agents studying L_b or L_c in countries other than X_a. Suppose not. Then $u_{ab} = rx_b$ and $u_{ac} = rx_c$. Since $x_b \neq x_c$ by definition, this contradicts $u_{ab} = u_{ac}$.

Let L_b be a language that a positive proportion of active agents in countries other than X_a studies. Consider a CC-deviation x_a of X_a to choose L_b. Then, by Lemma 1 (ii), an agent in this coalition gains a higher payoff than in the equilibrium. This is a contradiction. □

Proof of Proposition 2. **Situation 1.** Let $a, b, c, d \in N$. Situation 1 constitutes two subtypes. Situation 1-i is such that arcs construct a cyclic structure and in this structure, X_a is the head of the arc to X_b, X_b is that of X_c, X_c is that of X_d and X_d is that of X_a. Situation 1-ii is such that if X_a is the head of the arc to X_b, then X_a is the tail of the arc from X_b. See Figure 3 for an illustration. Any other graph does not belong to Situation 1. We suppose that those situations are equilibria and derive contradictions.

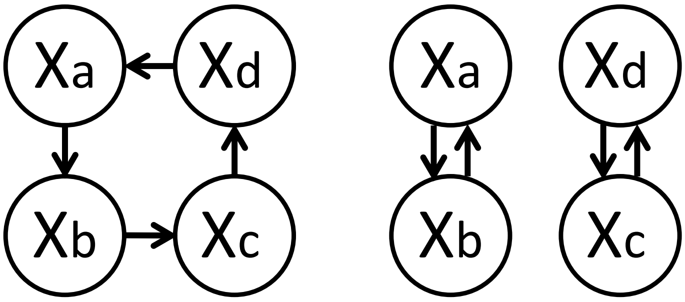

Figure 3. The left side illustrates Situation 1-i and the right side illustrates Situation 1-ii.

1-i. Suppose that Situation 1-i is an equilibrium. The gross payoff of agents in X_a in the equilibrium is

$$u_{ab}[x_a, 0, 0] = rx_b. \tag{10}$$

The payoff of agents in X_a CC-deviating to study L_d is

$$u_{ad}[x_a, x_c, 0] = \frac{r(1 + rx_c)x_d}{1 - r^2 x_a x_c}. \tag{11}$$

C1 implies that (10) $-$ (11) ≥ 0, which is

$$rx_b - \frac{r(1 + rx_c)x_d}{1 - r^2 x_a x_c} \geq 0 \iff \frac{x_b}{x_d} \geq \frac{1 + rx_c}{1 - r^2 x_a x_c}. \tag{12}$$

Similarly, the gross payoff of agents in X_c in the equilibrium is

$$u_{cd}[x_c, 0, 0] = rx_d. \tag{13}$$

The payoff of agents in X_c CC-deviating to study L_b is

$$u_{cb}[x_c, x_a, 0] = \frac{r(1 + rx_a)x_b}{1 - r^2 x_a x_c}. \tag{14}$$

C1 implies that (13) $-$ (14) ≥ 0, which is

$$rx_d - \frac{r(1 + rx_a)x_b}{1 - r^2 x_a x_c} \geq 0 \iff \frac{x_d}{x_b} \geq \frac{1 + rx_a}{1 - r^2 x_a x_c} \iff \frac{x_b}{x_d} \leq \frac{1 - r^2 x_a x_c}{1 + rx_a}. \tag{15}$$

Since $\{1 + rx_c\}/\{1 - r^2 x_a x_c\} > 1$, (12) implies that $x_b/x_d > 1$. Since $\{1 - r^2 x_a x_c\}/\{1 + rx_a\} < 1$, (15) implies that $x_b/x_d < 1$. These imply a contradiction.

1-ii. Suppose that Situation 1-ii is an equilibrium. As the X_b is the tail of the arc from X_a and X_d is that of X_c similar to 1-i, we can derive a contradiction in the same way.

Situation 2. The proof for this case is essentially the same as that for Situation 1. In Situation 2, there is a country that is the tails of the arcs from two countries. Let X_b be the heads of the arcs from X_a and X_c. Without loss of generality, let X_a be the head of the arc from X_d. Note that Situation 2 consists of Subsituations 2-i in which X_b is the tail of the arc to X_a, 2-ii in which X_b is the tail of the arc to X_c, and 2-iii in which X_b is the tail of the arc to X_d. (See Figure 4 for an illustration.) Suppose that Situation 2, which can be any of Subsituations 2-i, 2-ii and 2-iii at first, is an equilibrium, and we derive a contradiction.

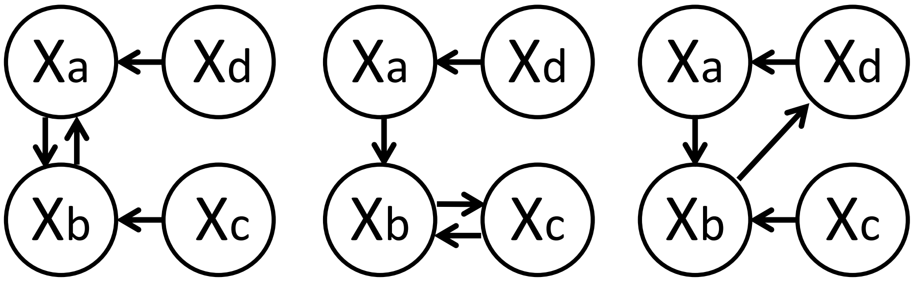

Figure 4. The left side illustrates Situation 2-i, the center illustrates 2-ii, and the right side illustrates 2-iii.

The gross payoff of agents in X_d in the equilibrium is

$$u_{da}[x_d, 0, 0] = rx_a. \tag{16}$$

The gross payoff of agents in X_d CC-deviating to study L_b is

$$u_{db}[x_d, x_a, x_c] = \frac{r(1 + rx_a)(1 + rx_c)x_b}{1 - r^2(x_a x_c + x_a x_d + x_c x_d) - 2r^3 x_a x_c x_d}. \tag{17}$$

C1 implies that (16) − (17) ≥ 0, which is

$$\begin{aligned} rx_a &\geq \frac{r(1 + rx_a)(1 + rx_c)x_b}{1 - r^2(x_a x_c + x_a x_d + x_c x_d) - 2r^3 x_a x_c x_d} \\ \Longleftrightarrow \quad \frac{x_a}{x_b} &\geq \frac{(1 + rx_a)(1 + rx_c)}{1 - r^2(x_a x_c + x_a x_d + x_c x_d) - 2r^3 x_a x_c x_d}. \end{aligned} \tag{18}$$

The gross payoff of agents in X_c in the equilibrium is

$$u_{cb}[x_c, x_a, 0] = \frac{r(1 + rx_a)x_b}{1 - r^2 x_a x_c}. \tag{19}$$

In Subsituations 2-ii and 2-iii, the gross payoff of agents in X_c CC-deviating to study L_a is

$$u_{ca}[x_c, x_d, 0] = \frac{r(1 + rx_d)x_a}{1 - r^2 x_c x_d}. \tag{20}$$

C1 implies that (19) − (20) ≥ 0, which is

$$\begin{aligned}
&\frac{r(1+rx_a)x_b}{1-r^2x_ax_c} - \frac{r(1+rx_d)x_a}{1-r^2x_cx_d} \geq 0 \\
\iff &\frac{(1+rx_a)(1-r^2x_cx_d)}{(1+rx_d)(1-r^2x_ax_c)} \geq \frac{x_a}{x_b} \\
\iff &\frac{(1+rx_a)(1-r^2x_cx_d)}{1+rx_d-r^2x_ax_c-r^3x_ax_cx_d} \geq \frac{x_a}{x_b}. \qquad (21)
\end{aligned}$$

Note that the left-hand side (LHS) of (18) equals the right-hand side (RHS) of (21). The denominator of the RHS of (18) is smaller than that of the LHS of (21), and the numerator of the RHS of (18) is larger than that of the LHS of (21). Thus, the RHS of (18) is larger than the LHS of (21). This fact, together with (18) and (21), implies a contradiction.

In Subsituation 2-i, the payoff of agents in X_c CC-deviating to study L_a is

$$u_{ca}[x_c, x_b, x_d] = \frac{r(1+rx_b)(1+rx_d)x_a}{1-r^2(x_bx_c+x_bx_d+x_cx_d)-2r^3x_bx_cx_d}. \qquad (22)$$

C1 implies that (19) − (22) ≥ 0, which is

$$\begin{aligned}
&\frac{r(1+rx_a)x_b}{1-r^2x_ax_c} - \frac{r(1+rx_b)(1+rx_d)x_a}{1-r^2x_bx_c-r^2x_bx_d-r^2x_cx_d-2r^3x_bx_cx_d} \geq 0 \\
\iff &\frac{(1+rx_a)x_b}{1-r^2x_ax_c} \geq \frac{(1+rx_b)(1+rx_d)x_a}{1-r^2x_bx_c-r^2x_bx_d-r^2x_cx_d-2r^3x_bx_cx_d} \\
\iff &\frac{(1+rx_a)(1-r^2x_bx_c-r^2x_bx_d-r^2x_cx_d-2r^3x_bx_cx_d)}{(1-r^2x_ax_c)(1+rx_b)(1+rx_d)} \geq \frac{x_a}{x_b}. \qquad (23)
\end{aligned}$$

The denominator of the RHS of (18) is smaller than that of the LHS of (23) and the numerator of the RHS of (18) is larger than that of the LHS of (23). Thus, the RHS of (18) is larger than the LHS of (23). This fact, together with (18) and (23), implies a contradiction. □