

広島大学学術情報リポジトリ  
Hiroshima University Institutional Repository

Title	Creación de un corpus oral para el estudio de la conversación en español de aprendientes japoneses de ELE
Author(s)	GARCÍA RUIZ-CASTILLO, Carlos
Citation	Hiroshima Studies in Language and Language Education , 25 : 155 - 170
Issue Date	2022-03-01
DOI	
Self DOI	
URL	<a href="https://ir.lib.hiroshima-u.ac.jp/00051967">https://ir.lib.hiroshima-u.ac.jp/00051967</a>
Right	Copyright (c) 2022 広島大学外国語教育研究センター
Relation	



# Creación de un corpus oral para el estudio de la conversación en español de aprendientes japoneses de ELE

Carlos GARCÍA RUIZ-CASTILLO

Instituto para la Investigación y la Enseñanza de Lenguas Extranjeras  
Universidad de Hiroshima

Las características de la interacción oral y de la conversación cotidiana pueden variar según la lengua y la cultura. Dado que existe una notable distancia lingüística entre el español y el japonés, el estilo conversacional en ambas lenguas también diferirá. Sin embargo, carecemos de análisis comparativos del estilo conversacional en estas lenguas. Asimismo, hasta el momento no se ha investigado la conversación en español como lengua extranjera (ELE) por parte de aprendientes cuya lengua materna es el japonés y, por tanto, no se dispone de un corpus adecuado para su estudio. En este trabajo presentamos nuestro corpus oral de conversaciones en ELE de aprendientes japoneses, justificamos nuestras decisiones en cuanto al diseño y recogida de datos y, mediante el análisis de dos extractos de conversaciones recogidas en el mismo, mostramos su adecuación para el estudio de las características del estilo conversacional de aprendientes japoneses de ELE.

## MARCO TÉORICO

### El estudio de la conversación y de la competencia interaccional

El Análisis de la Conversación (AC) concibe la interacción oral como un medio por el que dos o más participantes colaboran para llevar a cabo determinadas acciones sociales. El principal objetivo del AC es “describir, analizar y entender el hablar como una característica básica y constitutiva de la vida social” (Sidnell, 2010, p. 1)” mediante una perspectiva émica, esto es, sin categorías establecidas previamente por el investigador. El AC toma en consideración, entre otros, los siguientes sistemas de organización de la conversación interrelacionados entre sí: la toma de turnos (Sacks et al., 1974), la organización secuencial (Schegloff, 2007), la reparación (Hayashi et al., 2013), y las acciones que los participantes llevan a cabo e identifican en los turnos. Con una aproximación similar al AC pero centrada principalmente en analizar la influencia mutua entre las características de una determinada lengua y el modo de interactuar de los hablantes de esa lengua, se ha desarrollado en los últimos años la lingüística de la interacción (Couper-Kuhlen y Selting, 2018).

En el ámbito de la lingüística hispánica, destacan los trabajos del grupo Val.Es.Co sobre el español coloquial y la conversación cotidiana, y las investigaciones sobre la conversación con una perspectiva principalmente sociolingüística de Cestero Mancera (2000a, 2000b).

Dentro de la lingüística aplicada y la enseñanza de lenguas extranjeras, la interacción oral y la conversación han tenido diferentes concepciones a lo largo del tiempo (Pekarek Doehler, 2019), de forma que se han ido incorporando habilidades sociolingüísticas relacionadas con la conversación a la competencia comunicativa (Hymes, 1972), la competencia discursiva (Celce-Murcia et al., 1995) e incluso a la competencia conversacional (Celce-Murcia, 2007) o interaccional (Kasper, 2006). Estas diferentes concepciones inciden en que el objetivo del aprendizaje de una lengua extranjera (L2) es la capacidad de interactuar socialmente

con esa lengua. Los investigadores interesados en el desarrollo de la competencia interaccional de aprendientes de una L2 toman en consideración los sistemas antes mencionados de organización de la interacción oral procedentes del AC.

Disponemos de escasos estudios sobre la conversación en aprendientes de español como lengua extranjera (ELE). Destacan las investigaciones que se han realizado con aprendientes de ELE alemanes (García García, 2014) y taiwaneses (Pérez Ruiz, 2016), ambos trabajos con una metodología cuantitativa establecida por Cestero Mancera (2016). La descripción y el análisis de la conversación de aprendientes japoneses de ELE está por hacer.

### **La conversación en español como lengua extranjera de aprendientes japoneses**

La lingüística de la interacción asume que los sistemas de organización de la conversación descritos por el AC pueden ser universales y, al mismo tiempo, estar sujetos a variaciones según la lengua.

La construcción de los turnos de habla es un ejemplo de estas diferencias (Couper-Kuhlen y Selting, 2018). En las lenguas SVO, como el español, el elemento nuclear del enunciado, el verbo, se sitúa en una posición central. Los hablantes proyectan y marcan con señales lingüísticas la posible finalización del enunciado y la apertura, por ende, de un lugar pertinente para la alternancia de turno. Dado que los elementos que aparecen en el enunciado tras el verbo son informativamente más débiles, es frecuente que se den en esta posición interrupciones y habla solapada, característica importante de la conversación en español (Cestero Mancera, 2000a). Sin embargo, en las lenguas SOV, como el japonés, el verbo suele aparecer al final de los enunciados. Además, por las características aglutinantes de esta lengua, en los morfemas y partículas que aparecen junto al verbo se codifican significados esenciales, como la modalidad, el aspecto y el tiempo. De este modo, hasta el final de un enunciado los interlocutores no pueden tener una comprensión completa del mismo. Por ello, se ha considerado el japonés como una lengua con proyección tardía (Tanaka, 1999) y las superposiciones y las interrupciones serían menos frecuentes en esta lengua. También sería característica del japonés la alta frecuencia de turnos de apoyo o continuadores que emiten los interlocutores en espacios de microcolaboración (Iwasaki, 2009) durante el turno en marcha de un hablante.

Es decir, los diferentes estilos conversacionales en japonés y en español pueden estar influidos por aspectos lingüísticos. Es factible suponer que los aprendientes japoneses de ELE mostrarán rasgos de ambas lenguas en su conversación en ELE, y que estos rasgos variarán según el nivel de dominio. Para analizar las características del estilo conversacional de la interlengua de los aprendientes japoneses de ELE debemos obtener nuestros datos de conversaciones naturales. Sin embargo, el escaso corpus que recogen datos de estudiantes japoneses de ELE son corpus escritos (Valverde, 2020) y no hemos localizado ningún corpus oral formado por conversaciones. En definitiva, resulta necesaria la creación de un corpus oral para el estudio de la conversación en ELE de aprendientes japoneses.

### **CREACIÓN DE UN CORPUS ORAL DE CONVERSACIONES DE APRENDIENTES JAPONESES DE ELE**

En este estudio preparatorio pretendemos justificar las decisiones que guían el diseño y la creación de nuestro corpus oral de conversaciones en ELE de aprendientes japoneses. Además, queremos comprobar si nuestro corpus permite obtener datos relevantes para caracterizar el estilo conversacional de estos aprendientes.

Para ello, responderemos a las siguientes preguntas de investigación:

1. ¿Cómo diseñar un corpus oral que permita analizar y describir las conversaciones en ELE de aprendientes japoneses?
2. ¿Cómo recoger los datos del corpus y tratarlos para el análisis y descripción de las conversaciones en ELE de aprendientes japoneses?
3. ¿Un corpus oral de conversaciones puede ofrecer datos relevantes para caracterizar el estilo conversacional y la competencia interaccional de los aprendientes japoneses de ELE?

### **Diseño del corpus**

Este corpus está dirigido principalmente a investigadores de la interacción oral como actividad social por aprendientes de una L2. El corpus también puede ser de interés para profesores de ELE con estudiantes cuya lengua materna sea el japonés y, en última instancia, para aprendientes japoneses de ELE.

De acuerdo con nuestro interés investigador, el corpus recogerá interacciones de aprendientes japoneses de ELE agrupadas en dos niveles diferentes de dominio: tras dos años de estudio de la L2 en programas de estudios universitarios de grado de español como especialidad (grupo A) y tras cuatro años (grupo B), respectivamente. Se ha considerado de manera aproximada que tras dos años estudios los estudiantes alcanzan un nivel A2 y, tras cuatro años, entre B1 y B2.

Nuestro corpus de aprendientes debe permitir análisis comparativos con los corpus orales y datos sobre la conversación disponibles para el español y el japonés como lenguas maternas. Esperamos realizar estudios comparativos tomando como referencia, al menos, los datos ofrecidos por Cestero Mancera (2000a, 2000b) y el corpus de conversaciones en japonés BTSJ (Usami, 2021).

En el diseño de nuestro corpus seguimos las fases habituales de una investigación en el ámbito del AC: grabación de interacciones naturales, transcripción de las grabaciones en su totalidad o en parte, análisis de episodios seleccionados según el objeto de estudio y, finalmente, comunicación de los resultados (Ten Have, 2007).

En definitiva, creamos nuestro corpus con la finalidad de obtener datos que permitan la descripción del estilo conversacional de los participantes durante su interacción oral. No tiene, por tanto, naturaleza experimental. Finalmente, nuestro corpus no está concebido en un principio como un corpus de análisis lingüístico del léxico o de aspectos gramaticales, por lo que no incluye etiquetado estructural o gramatical.

### Crterios para la recogida de datos

El corpus está formado exclusivamente por conversaciones cotidianas mantenidas por aprendientes japoneses de ELE. En base a tres rasgos definitorios, caracterizamos la conversación cotidiana como un tipo de interacción oral con un bajo grado de convencionalización y planificación, sin un objetivo social predeterminado salvo el establecimiento y mantenimiento de relaciones interpersonales y en la que ninguno de los participantes, que tienen la misma categoría real o funcional, controla individualmente la distribución de los turnos (Cestero Mancera, 2005).

En relación con el procedimiento de obtención de datos, las conversaciones de nuestro corpus son semidirigidas (Moreno Fernández, 2009), por dos razones: la toma de turnos no está establecida con antelación, como en una entrevista, y las conversaciones están planificadas por el investigador con unos

temas o propósitos específicos, al contrario que en las conversaciones libres. El control del investigador en nuestro corpus consiste en ofrecer a los participantes una lista de posibles temas a desarrollar durante la grabación, con la opción de elegir otros temas libremente. El investigador no interviene en las grabaciones. Finalmente, el tipo de grabación es ordinaria, no secreta, con los dispositivos de grabación a la vista de los participantes.

Seguimos para la elección de nuestros temas la metodología del proyecto PRESEEA (PRESEEA, 2003), según la cual es conveniente proponer temas cercanos a la realidad inmediata de los participantes y que favorezcan la aparición de discursos más espontáneos. Seleccionamos temas adecuados para los dos niveles de dominio, tales como su vida cotidiana, su experiencia como estudiantes de ELE, acontecimientos importantes en su vida, la importancia del aprendizaje de ELE en su futuro, etc.

Estimamos que una transcripción de 10 minutos de cada grabación es suficiente para obtener ejemplos relevantes de los fenómenos conversacionales que se desee estudiar. Se contempla grabar una duración mayor en cada conversación dado que en los primeros minutos los participantes suelen mostrar cierto nerviosismo y un habla más cuidada (Moreno Fernández, 2009), y para no cortar la secuencia temática en curso al final de la conversación.

#### Participantes y tamaño del corpus

Para nuestro corpus de conversaciones nos centramos en un contexto específico, los estudios universitarios de español como especialidad en universidades japonesas que cuentan con Departamento de Español o Departamento de Estudios Hispánicos. Elegimos este contexto por la relativa uniformidad de los informantes y facilidad de acceso.

Los programas de estos departamentos suelen incluir 5 clases (es decir, 7,5 horas) o más de ELE a la semana en cursos de 30 semanas. Las clases de ELE suelen ser obligatorias en los dos primeros años, mientras que en los dos últimos los programas se enfocan en contenidos específicos o de área. Por ello, el nivel de dominio de la L2 es generalmente A2 al finalizar los dos primeros años, mientras que puede ser muy dispar al finalizar los cuatro años de estudios (Valverde, 2020). Los estudios de español como especialidad se ofrecen en 15 universidades japonesas (Ugarte Ferrerons, 2012) y los siguen unos 4500 estudiantes (Sanz Yagüe, 2013). Así pues, consideramos que nuestra población es 4500 individuos.

Establecimos dos grupos (A y B) según el nivel de dominio de la L2 y empleamos un muestreo selectivo por cuotas de afijación uniforme. Para completar las cuotas de 30 participantes por grupo seguimos un procedimiento similar al muestreo accidental, basado en la disponibilidad y voluntariedad de los informantes a participar en las grabaciones. La muestra de nuestro corpus incluye 60 informantes, lo que supera con creces el criterio habitual en la sociolingüística variacionista (0,025%). Esto se debe, por un lado, a que el objetivo principal de nuestro corpus es recoger un número de interacciones suficiente como para obtener ejemplos del fenómeno conversacional que se desee estudiar, y por otro, a que los fenómenos conversacionales presentan una gran variación según las relaciones entre los participantes, los temas tratados y otros factores.

En la siguiente tabla recogemos los criterios de inclusión de los participantes por cada grupo del corpus atendiendo a factores sociológicos y lingüísticos.

**TABLA 1. Criterios de inclusión de participantes**

	Grupo A	Grupo B
<b>Número de participantes</b>	30	30
<b>Factores sociológicos</b>		
Edad	19-20 años	21-23 años
Nivel educativo	universidad	universidad
Lugar de estudio	Departamento de Español o de Estudios Hispánicos	Departamento de Español o de Estudios Hispánicos
<b>Factores lingüísticos</b>		
Lengua materna	japonés	japonés
Lengua meta	español	español
Curso del programa de especialidad	final del 2º curso	final del 4º curso
Nivel de español esperado	A2	B1-B2

### **Recogida de datos y su tratamiento**

Para realizar las grabaciones, acudimos a cuatro universidades japonesas con programas de español o de estudios hispánicos como especialidad: la Universidad Sofía, la Universidad de Estudios Extranjeros de Tokio, la Universidad de Estudios Extranjeros de Kioto y la Universidad Municipal de Estudios Extranjeros de Kobe. Según se detalla en el apéndice 2, las grabaciones tuvieron lugar en diversos momentos entre finales de 2018 y finales de 2019, en los meses correspondientes al segundo semestre del año académico en Japón. Se contactó con los participantes gracias a la colaboración de los profesores de dichas universidades o a la relación directa del investigador con los estudiantes por haber impartido clase de ELE en alguna de esas instituciones.

Se solicitó a los participantes que completaran dos formularios: uno relativo al consentimiento informado para el uso académico de las grabaciones en audio y vídeo y sus transcripciones, y otro con datos sociolingüísticos (edad, sexo, lengua materna, universidad, curso en la universidad, año de ingreso en la universidad, tiempo de estudio del español y tiempo de estancia en países hispanohablantes).

No realizamos pruebas previas para obtener el nivel preciso de dominio de la L2. Como hemos señalado anteriormente, al finalizar los dos primeros años del programa de estudios es relativamente uniforme. En el caso del grupo B hay gran disparidad. Por ello, nos guiamos principalmente por la valoración de los profesores de las respectivas universidades. Algunas de las grabaciones se desestimaron por una baja fluidez que dificultaba el desarrollo de la conversación.

### Perfil sociolingüístico de los participantes

Además de los criterios de inclusión de los participantes y del criterio de afijación de las cuotas, tomamos otros datos que pueden ser relevantes en nuestro estudio.

En cuanto al sexo, hacemos notar que en el corpus el número de hombres y mujeres y de conversaciones entre participantes del mismo sexo y de diferente sexo es desigual.

Por lo que respecta al tiempo de estudio de ELE, en el grupo A va desde los 18 meses hasta los 21, y solo un informante había estudiado ELE antes de entrar en la universidad. Casi todos los informantes del grupo B cursaban el cuarto curso de sus programas, si bien 5 estaban en el quinto año de los estudios universitarios. Hacemos notar que, por indicación de sus profesores, aceptamos la participación de un

estudiante del tercer curso, dado que su nivel de dominio era comparable al de cursos superiores.

La mayor parte de los estudiantes del grupo A no había estado en países hispanohablantes o habían realizado estancias cortas. Sin embargo, la mayoría de los informantes del grupo B había realizado estancias prolongadas de estudios o prácticas laborales en países hispanohablantes.

En la siguiente tabla incluimos un resumen de los datos expuestos hasta ahora.

**TABLA 2. Perfil sociolingüístico de los participantes**

	Grupo A	Grupo B
Número de participantes	30	30
Número de conversaciones	15	15
Participantes femeninos	21	23
Participantes masculinos	9	7
Conversaciones entre dos participantes femeninos	10	5
Conversaciones entre dos participantes masculinos	4	2
Conversaciones entre participantes de diferente sexo	1	8
Promedio de tiempo de estudio de ELE en meses	20,2	45,5
Promedio de tiempo de estancia en países hispanohablantes en meses	0,45	8

#### Aspectos técnicos de las grabaciones

Las grabaciones se realizaron en las respectivas universidades, en salas con el menor ruido ambiente posible y en las que solo estuvieran los participantes y el investigador. Frente a los dos participantes y en una posición equidistante a ambos se situaron dos grabadoras de sonido, una de ellas de seguridad o respaldo. Se dispusieron dos cámaras de vídeo, cada una dirigida a uno de los participantes. Para las grabaciones de audio se usaron dos grabadoras Zoom H2N y, para el vídeo, una cámara Panasonic Lumix DMC-LX100 y una cámara Sony A7. Empleamos los formatos WAV y MP4. Tras las grabaciones, se ajustó el tamaño del vídeo de cada participante para formar uno único con ambas tomas y se sincronizó la imagen con el audio de una de las grabadoras.

Señalamos que las grabaciones de audio no se recogieron en canales diferentes para cada participante, lo cual puede dificultar las investigaciones en las que se requiera calcular con precisión fenómenos prosódicos como la intensidad o el tono.

#### Transcripción de los materiales

Una vez preparados los materiales audiovisuales, procedimos a su transcripción. De cada conversación transcribimos diez minutos. Evitamos en lo posible los primeros minutos y cortar secuencias temáticas en curso.

Para la transcripción usamos el programa ELAN (Max Planck Institute for Psycholinguistics, The Language Archive, 2021). ELAN es un programa para el análisis y transcripción de materiales audiovisuales. Las anotaciones se registran en múltiples niveles o líneas dispuestas en el eje temporal, en un estilo denominado “de partitura” (Hepburn y Bolden, 2017). Además, muestra tanto el oscilograma como el contorno melódico, permite intercambiar datos fácilmente con el programa de análisis del habla PRAAT (Boersma y Weenink, 2021), y admite la importación y la exportación en formatos estándares, tales como

CSV, XML o HTML. Se trata, en suma, de un programa adecuado para la transcripción de la conducta social y aspectos lingüísticos y multimodales (Hepburn y Bolden, 2017).

ELAN no admite, sin embargo, la transcripción según el sistema y convenciones creados por Gail Jefferson y usados comúnmente en el AC (Hepburn y Bolden, 2017). Por ello, usamos un sistema de transcripción propio basado principalmente en los trabajos del grupo Val.Es.Co (Briz Gómez, 2016) y del proyecto PRESEEA (PRESEEA, 2008). Seguimos en nuestro sistema los requisitos que señala Briz Gómez (2016) para un sistema de transcripción: que cada signo represente un único fenómeno (exhaustividad), que cada fenómeno aparezca codificado por un único signo (pertinencia), que sea posible aumentar o reducir el número de signos que lo forman (adaptabilidad) y que la transcripción resultante pueda ser revisada y verificada (validez). Además, coincidimos con este autor en la necesidad de que el sistema de transcripción refleje fielmente y, al mismo tiempo sin impedir la lectura fluida, fenómenos conversacionales como alternancia de turnos, solapamientos, etc. En el apéndice 1 recogemos las convenciones seguidas en nuestro sistema de transcripción.

#### Precisiones en cuanto a nuestro sistema de transcripción

En nuestro sistema de transcripción mediante ELAN, creamos una línea por cada hablante y segmentamos cada línea en el eje temporal. En cada anotación recogemos la transcripción del mensaje lingüístico y ciertos fenómenos paralingüísticos como la risa. No transcribimos, por el momento, gestos o elementos quinésicos y próxemicos.

Los criterios para la segmentación se basan en las cesuras o pausas y el intercambio de hablante. Es decir, si se da una alternancia de turno o una pausa en el turno del hablante, generalmente de larga duración, creamos una nueva anotación. Por ello, en nuestro sistema, las anotaciones no se corresponden con unidades o subunidades del discurso previamente establecidas.

En cuanto la entonación, recogemos la inflexión melódica ascendente. Nos guiamos principalmente por criterios de percepción, si bien acudimos a PRAAT para confirmarlo. La inflexión melódica ascendente, al igual que otras marcas como los alargamientos, son recursos muy usados por los aprendientes japoneses de ELE para indicar la continuación del turno en marcha. No señalamos la inflexión melódica al final de las preguntas, pero sí en su interior.

Por lo que respecta a la transcripción de la duración de fenómenos prosódicos y conversacionales, como el alargamiento de sonidos o la risa, encontramos dos posturas al respecto. En el sistema creado por Gail Jefferson se opta por representar mediante la longitud de los signos o su repetición la duración del fenómeno. En los sistemas de transcripción usados en la conversación en español esta representación es muy poco precisa y se suele basar en la mera repetición del signo para indicar una mayor duración. En nuestro caso, dado que ELAN no permite usar adecuadamente las convenciones de transcripción del AC, preferimos indicar solamente la presencia del fenómeno, pero no su duración.

En la actualidad hemos completado la transcripción, con el procedimiento y nivel de detalle ya señalados de aproximadamente 10 minutos de cada una de las 30 conversaciones; es decir, 302 minutos en total de los 409 minutos disponibles. En el apéndice 2 indicamos la duración total de cada conversación y el intervalo descrito.

Hemos señalado que uno de los requisitos de un sistema de transcripción es que sea válido. Para ello,



debe ser revisado y verificado mediante filtros internos y externos (Briz Gómez, 2016). Por el momento, nuestro corpus solo ha sido sometido al filtro interno del propio investigador, por lo que aún debe ser revisado por investigadores externos.

Extractos de dos conversaciones recogidas en nuestro corpus

Para ilustrar lo expuesto hasta el momento, recogemos dos ejemplos de conversaciones incluidas en nuestro corpus y realizamos un primer análisis secuencial.

El primer extracto pertenece a la conversación etiquetada como A01, grabada en la Universidad Sofia. Las dos participantes (01M y 02M) son estudiantes de segundo curso y mujeres. Recogemos las anotaciones 187-206, correspondientes al intervalo 10:12-10:43 de la grabación. En los intercambios previos, no recogidos en este extracto, 01M transmite su deseo de ser diplomática en el futuro a 02M y en la anotación 187 justifica su deseo, tras lo cual 02M la elogia en 194. Esto lleva a 01M a restar importancia a su aspiración y a reconocer las dificultades que conllevará realizar su sueño (195-202). En las anotaciones 203-206 aparecen varios turnos breves que cierran la secuencia.

EXTRACTO 1. Conversación A01.

```

187 01M quiero(.) [tra]bajar como diplomáti[ca] porque (1.3) ha[y-] me(:)
      interesa mucho internacional
188 02M          [(a)]
189          [(a)]
190          [(•h)]
      (0.3)
191          [°(eh)°]
192 01M [in-]/ relaciones internacionale[s]
193 02M          [(a)]
      (0.9)
194 te queda MUY bien
195 01M (risa) fnoɸ (r[is]a) [es d-]/ [es-]/ tengo que estudiar mucho
      por[que]el examen es muy di[fi]cilt†
196 02M          [&ja&]
197          [(( ))]
198          [(m)]
199          [(m)]
200          [(a)]
      (0.6)
201 (a)/ difíci- (a:)/ exámenes/ in[tenta]
202 01M          [para(:)] para s- para se(:)r(.)
      diplomática
      (0.2)
203 02M °i(a:)!°
      (0.3)
204 °sí°
      (0.1)
205 01M (ri[sa])
206 02M [(risa)] (a)

```

El extracto 2 corresponde al intervalo 5:56-7:15 de la conversación B11. Las dos participantes (51M, 52M), en este caso estudiantes de cuarto año y también mujeres, hablan de la estancia de 52M en México. El número de horas que allí trabajaba 52M sorprende a 51M (111). Se siguen diferentes explicaciones de 52M (112-116). En 117 y 120, 51M da por concluida esta secuencia; sin embargo, no emite ningún turno nuevo, por lo que 052M expande el suyo hasta en dos ocasiones (118-119, 122-125). Tras unos turnos breves y un silencio de 2 segundos, 51M inicia una nueva secuencia sobre su estancia en Valencia.

EXTRACTO 2. Conversación B11.

107 51M [(ah)] // ¿cuántos(:)// cuántas horas(.) trabajas  
(0.4)

108 52M ¿allá[:)]?

109 51M [al] día?  
(0.1)

110 52M allá (1,3) al día(:)// °(ah)°(.) nueve horas o más o menos creo  
(0.8)

111 51M [¿nueve horas?]

112 52M [es que siem-] siempre siempre empezamos(.) empezamos a(:)(.)  
la(:)s(:)// a las ocho†/ hasta(:)// hasta cinco(.) cinco y media//  
con(:) tre-/ con treinta minutos de[:)]

113 51M [¿descan]so?

114 52M sí de des[canso]

115 51M [¿solo?]/ ¿solo [treinta minutos?]

116 52M [nada más](.) nada más treinta minutos(.)  
por eso siempre siempre tenemos(.) teníamos que comer muy/ rápido  
(0.5)

117 51M (ah:)  
(0.6)

118 52M (risa)  
(1.2)

119 pero bueno// mientras mientras† tra-/ mientras trabajando(.) también  
descansamos poquito a poquito  
(0.1)

120 51M (m[:]) (sí (.)) sí sí]

121 52M [(risa)]  
(1.4)

122 por eso/ [en total†]/ ocho/ ocho y cinco y media (1,8) (eh)

123 51M [(risa)]  
(0.2)

124 ¿(m)?/ ¿ocho horas y/ [media?]

125 52M [cua]tro/ cuatro más cinco/ i(ah)! bueno(.)  
nueve- nueve(.) horas  
(1.0)

126 51M es mucho=

127 52M es mucho(.) [es] mucho [(ri)sa)

128 51M [sí]

129 [sí]  
(2.0)

130 [cuando(:)]/ cuando yo estaba en España↑// no/ no tenía qué hacer  
 (1.9) siempre dormía

131 52M [°(m:)°]  
 (0.0)

132 (ri[sa])

133 51M [(risa)]// siesta(.) [siesta] siesta

134 52M [pero sa-]  
 (0.2)

135 pero también saliste muchas veces ¿no?// es que yo vi [...]

## DISCUSIÓN

En los apartados previos hemos establecido una perspectiva basada en el AC, la competencia interaccional en L2 y la lingüística de la interacción con la que realizar nuestra investigación sobre el estilo conversacional de los aprendientes japoneses de ELE. Asimismo, hemos justificado las decisiones de diseño y recogida de materiales de nuestro corpus. En este apartado analizaremos los extractos 1 y 2 incluidos como ejemplo. En ellos podemos observar diferencias que quizás reflejen la influencia del nivel de dominio y de la lengua materna en el estilo conversacional de los participantes.

En el extracto 1, la participante 01M diseña cada uno de sus turnos 187 y 195 con al menos dos unidades constitutivas de turno y las pausas interiores, reinicios y dubitaciones propias del nivel de dominio. Observamos cómo 02M sitúa sus turnos de apoyo o continuadores superpuestos con precisión en dos momentos estratégicos de los turnos de 01M: tras pausa y, por tanto, en el inicio de un nuevo grupo entonativo (188, 190, 198), y en posición final del grupo entonativo (189, 200). Es decir, las disfluencias en la construcción del turno se corresponden con una alta frecuencia de turnos de apoyo o continuadores en el interior del turno en marcha. Esta frecuencia y posición de los turnos de apoyo es similar a la que se ha descrito para el japonés. Observamos también cómo 52M usa determinadas vocalizaciones para marcar el inicio de un nuevo turno (191 y 193), así como una vocalización de mayor duración y con una entonación enfática (203) que parece indicar entendimiento y final de secuencia. Las participantes intervienen en el final de la secuencia con turnos breves y elementos paralingüísticos como la risa; son postexpansiones mínimas con la doble función de emitir un turno en un momento requerido y, al mismo tiempo, indicar que no se abre una nueva secuencia.

En contraste, en el extracto 2 observamos fenómenos diferentes, especialmente en cuanto a la organización secuencial y al diseño de los turnos, que son más complejos, presentan menos disfluencias y mayor integración de los grupos entonativos. Apenas encontramos turnos de apoyo o continuadores en el interior de los turnos ni vocalizaciones para marcar el inicio de turno. Este fragmento es, además, rico en superposiciones. Son relevantes porque el español es una lengua que destaca por la alta frecuencia del habla superpuesta, y además muestran cómo los participantes organizan las alternancias y registran los turnos de los interlocutores atentos a la proyección de los posibles lugares de transición relevantes. Encontramos, por ejemplo, superposiciones en posición inicial de turno (111-112 y 133-134). Apreciamos también superposiciones de comprensión al final de turno, al quedar este proyectado y claro para el interlocutor (112-113, 114-115, 115-116 y 124-125). En la superposición en 113 encontramos un caso de compleción del turno por el interlocutor; es decir, 51M completa sintácticamente el turno de 52 ante sus dubitaciones y

alargamientos. Finalmente, en el extracto 2 observamos recursos diferentes a los del extracto 1 para manejar el final de secuencia. Tras pausas, vocalizaciones y elementos paralingüísticos (117-119, 120-121), 52M inicia un nuevo turno (119 y 122) que expande la secuencia; observamos que tiene éxito en el segundo caso, puesto que involucra a 51M, si bien mediante reparación. Tras una pausa de cierta duración, en 130 la participante 52M abre una nueva secuencia.

## CONCLUSIONES

En el presente trabajo hemos justificado la necesidad de comprender cómo los aprendientes japoneses de ELE llevan a cabo una conversación cotidiana como medio de interacción social en la L2. Dado que nuestro interés principal es analizar si la lengua materna y el nivel de dominio influyen en las características de la interacción oral de estos aprendientes, hemos creado un corpus oral de conversaciones mantenidas por aprendientes japoneses con dos niveles de dominio de ELE diferentes. Hemos descrito y justificado las decisiones en cuanto al diseño y recogida y tratamiento de materiales de nuestro corpus.

En los dos extractos analizados hemos encontrado diferencias notables en el estilo conversacional de los aprendientes. La frecuencia y posición de los turnos de apoyo parecen acercar el estilo conversacional de los aprendientes con menor nivel de dominio de ELE al del japonés como lengua materna, mientras que la frecuencia y tipos de habla solapada en los aprendientes con un nivel de dominio más alto quizás se relacione con el estilo conversacional del español como lengua materna. Este breve análisis preliminar ha mostrado que nuestro corpus puede servir para el estudio de la conversación y de la competencia interaccional de aprendientes japoneses de ELE. En posteriores estudios debemos confirmar y comprobar con más detalle estas consideraciones.

## REFERENCIAS

- Boersma, P., y Weenink, D. (2021). *Praat: Doing phonetics by computer* (6.2.01) [Software informático]. <http://www.praat.org/>
- Briz Gómez, A. (2016). El proyecto AMERESCO. La idea de un corpus de conversaciones coloquiales del español de América. En A. M. Bañón Hernández, M. del M. Espejo Muriel, B. Herrero Muñoz-Cobo, y J. L. López Cruces (Eds.), *Oralidad y análisis del discurso. Homenaje a Luis Cortés Rodríguez* (pp. 81-104). Editorial Universidad de Almería.
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language Teaching. En E. Alcón Soler y M. P. Safont Jordá (Eds.), *Intercultural language use and language learning* (pp. 7-22). Springer.
- Celce-Murcia, M., Dörnyei, Z., y Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied linguistics*, 6(2), 5-35. <https://doi.org/10.5070/L462005216>
- Cestero Mancera, A. M. (2000a). *El intercambio de turnos de habla en la conversación (análisis sociolingüístico)*. Universidad de Alcalá.
- Cestero Mancera, A. M. (2000b). *Los turnos de apoyo conversacionales*. Universidad de Cádiz, Servicio de Publicaciones.
- Cestero Mancera, A. M. (2005). *Conversación y enseñanza de lenguas extranjeras*. Arco Libros.

- Cestero Mancera, A. M. (2016). La conversación. Bases teóricas y metodológicas para su investigación y su enseñanza en ELE. *Linred: Lingüística en la red*, XIV.
- Couper-Kuhlen, E., y Selting, M. (2018). *Interactional linguistics: Studying language in social interaction*. Cambridge University Press.
- García García, M. (2014). *La competencia conversacional en español como lengua extranjera. Análisis y enfoque didáctico*. Universidad de Alcalá.
- Hayashi, M., Raymond, G., y Sidnell, J. (Eds.). (2013). *Conversational repair and human understanding*. Cambridge University Press.
- Hepburn, A., y Bolden, G. B. (2017). *Transcribing for social research*. SAGE Publications.
- Hymes, D. (1972). On communicative competence. En J. B. Pride y J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269-293). Penguin.
- Iwasaki, S. (2009). Initiating interactive turn spaces in Japanese conversation: Local projection and collaborative action. *Discourse Processes*, 46(2-3), 226-246.
- Kasper, G. (2006). Beyond repair. *AILA Review*, 19(1), 83-99. <https://doi.org/10.1075/aila.19.07kas>
- Max Planck Institute for Psycholinguistics, The Language Archive. (2021). ELAN (6.2) [Software informático]. <https://archive.mpi.nl/tla/elan>
- Moreno Fernández, F. (2009). *Principios de sociolingüística y sociología del lenguaje*. Ariel.
- Pekarek Doehler, S. (2019). On the nature and the development of L2 interactional competence. En M. R. Salaberry y S. Kunitz (Eds.), *Teaching and testing L2 interactional competence* (pp. 52-103). Routledge.
- Pérez Ruiz, J. (2016). Los turnos de apoyo verbales en las conversaciones españolas de estudiantes taiwaneses de ELE. *Lingüística en la Red*, XIV
- PRESEEA. (2003). Metodología del “Proyecto para el estudio sociolingüístico del español de España y América (PRESEEA)”. <http://preseea.linguas.net/Portals/0/Metodologia/METODOLOGÍA%20PRESEEA.pdf>
- PRESEEA. (2008). Marcas y etiquetas mínimas obligatorias. [http://preseea.linguas.net/Portals/0/Metodologia/Marcas\\_etiquetas\\_minimas\\_obligatorias\\_1\\_2.pdf](http://preseea.linguas.net/Portals/0/Metodologia/Marcas_etiquetas_minimas_obligatorias_1_2.pdf)
- Sacks, H., Schegloff, E. A., y Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735. <https://doi.org/10.2307/412243>
- Sanz Yagüe, M. (2013). Situación presente y futura del español en Japón. Elucubraciones de una profesora dirigidas a otros profesores en Asia (y a las autoridades). <http://languagescience-kobe.org/academic-blog/2/21/2013/-situacin-presente-y-futura-del-espao1-en-japn-elucubraciones-de-una-profesora-dirigidas-a-otros-profesores-en-asia-y-a-las-autoridades>
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.
- Sidnell, J. (2010). *Conversation analysis: An introduction*. Willey-Blackwell.
- Tanaka, H. (1999). *Turn-taking in Japanese conversation: A study in grammar and interaction*. John Benjamins.
- Ten Have, P. (2007). *Doing conversation analysis*. SAGE Publications.
- Ugarte Ferrerons, V. (2012). El español en Japón. En *El español en el mundo. Anuario del Instituto Cervantes. Instituto Cervantes*. [https://cvc.cervantes.es/lengua/anuario/anuario\\_12/ugarte/p01.htm](https://cvc.cervantes.es/lengua/anuario/anuario_12/ugarte/p01.htm)
- Usami, M. (2021). *BTSJ-Japanese Natural Conversation Corpus with Transcripts and Recordings (March,*

2021). NINJAL Institute-based projects: Multiple Approaches to Analyzing the Communication of Japanese Language Learners.

Valverde, P. (2020). Diseño y creación de un corpus de aprendices de español en Japón (CELEN). *E-AESLA*, 6, 223-240.

## APÉNDICE 1. CONVENCIONES DE TRANSCRIPCIÓN

[ ]	Inicio y final de habla solapada.
-	Reinicio y autointerrupción sin pausa.
(.)	Pausa corta, difícil de medir o inferior a 0,2 segundos.
/	Pausa inferior a 0,5 segundos.
//	Pausa entre medio segundo y un segundo.
(0.7)	Indicación en segundos de la duración de una pausa, lapso o silencio.
(:)	Alargamiento del sonido al que precede.
(h)	Espiración audible y con posible implicación secuencial.
(•h)	Aspiración audible y con posible implicación secuencial.
£ £	Discurso emitido con voz sonriente.
(risa, tos...)	Fenómenos relacionados con la interacción y la emisión del turno.
(( ))	Transcripción dudosa. Si aparece en blanco, porción ininteligible.
° °	Porción de discurso emitida con intensidad baja, próxima al susurro.
& &	Porción de discurso en otro idioma. Si se trata de japonés, se ha transliterado.
(.pk)	Chasquido de lengua
(m), ¡(a)!	Vocalizaciones
¡!	Exclamaciones
¿?	Interrogaciones y apéndices del tipo “¿no?”.
↑	Inflexión melódica ascendente.
=	Sucesión inmediata de turno.

## APÉNDICE 2. DATOS DE LAS GRABACIONES

Código conversación	Fecha de la grabación	Universidad	Tiempo de grabación	Inicio del intervalo de transcripción	Final del intervalo de transcripción	Tiempo de transcripción
A01	14/12/2018	Sofía	15:04	4:52:00	14:52:00	10:00
A02	14/12/2018	Sofía	18:14	2:47:00	12:56:00	10:09
A03	19/12/2018	Sofía	13:15	2:30:00	12:39:00	10:09
A04	19/12/2018	Sofía	11:35	1:11:00	11:16:00	10:05
A05	20/12/2018	Sofía	14:38	3:02:00	13:22:00	10:20
A06	20/01/2019	Sofía	14:16	4:19:00	14:06:00	9:47
A07	10/01/2019	TUFS	15:19	2:42:00	12:32:00	9:50
A08	10/01/2019	TUFS	15:53	1:40:00	11:44:00	10:04
A09	14/10/2019	TUFS	13:22	2:47:00	12:55:00	10:08
A10	14/10/2019	TUFS	15:26	4:48:00	14:51:00	10:03
A11	14/10/2019	TUFS	12:57	1:45:00	11:47:00	10:02
A12	14/10/2019	TUFS	16:14	2:50:00	12:54:00	10:04
A13	25/11/2019	KUFS	15:18	2:47:00	13:07:00	10:20
A14	25/11/2019	KUFS	12:33	1:17:00	11:22:00	10:05
A15	25/11/2019	KUFS	15:42	5:16:00	15:20:00	10:04
B01	14/10/2019	TUFS	14:03	4:00:00	14:01:00	10:01
B02	14/10/2019	TUFS	12:46	2:14:00	12:25:00	10:11
B03	06/11/2019	Sofía	14:01	3:31:00	13:38:00	10:07
B04	12/12/2019	KUFS	13:13	1:17:00	11:22:00	10:05
B05	06/11/2019	Sofía	13:36	0:07:00	10:39:00	10:32
B06	21/12/2018	TUFS	11:39	1:13:00	11:13:00	10:00
B07	25/11/2019	KUFS	13:47	3:04:00	12:54:00	9:50
B08	12/12/2019	KCUFS	12:10	1:43:00	11:49:00	10:06
B09	12/12/2019	KCUFS	12:19	2:21:00	12:14:00	9:53
B10	12/12/2019	KCUFS	11:28	0:43:00	10:47:00	10:04
B11	12/12/2019	KCUFS	12:01	1:34:00	11:53:00	10:19
B12	12/12/2019	KCUFS	11:43	1:21:00	11:21:00	10:00
B13	12/12/2019	KCUFS	11:27	1:18:00	11:16:00	9:58
B14	12/12/2019	KCUFS	11:18	1:13:00	11:09:00	9:56
B15	12/12/2019	KCUFS	14:01	2:44:00	12:44:00	10:00

*Nota:* Sofía (Universidad Sofía, Tokio), TUFS (Universidad de Estudios Extranjeros de Tokio), KUFS (Universidad de Estudios Extranjeros de Kioto), KCUFS (Universidad Municipal de Estudios Extranjeros Kobe)

## ABSTRACT

### **Creation of an Oral Corpus for the Study of Japanese Students' Conversation in Spanish as a Foreign Language**

Carlos GARCÍA RUIZ-CASTILLO

Institute for Foreign Language Research and Education  
Hiroshima University

Despite the importance of oral interaction and conversation in foreign language learning, there is a striking lack of studies regarding the everyday conversation of Japanese students in Spanish as a Foreign Language (SFL). As Japanese and Spanish are two languages with a remarkable linguistic distance, it is reasonable to suppose that the conversational style of Japanese learners of SFL will be different from the style of speakers of either Japanese or Spanish as a mother tongue, and it will differ across competence levels.

In this paper, I present the first oral corpus created to research the conversational style of Japanese Learners of SFL with a methodology based on Conversation Analysis and Interactional Linguistics. First, I discuss the protocols for designing the corpus and the procedures for collecting the data. Second, I present two extracts as examples of the conversations included in the corpus. Finally, I analyze these two extracts. This preliminary analysis suggests that participants with a lower competence level may use a conversational style closer to Japanese, with more continuers or backchannels during the ongoing turn, while higher level participants produce fewer continuers and more overlapping talk at the end of the turns, a characteristic feature of conversation in Spanish.



## 要 約

### スペイン語会話研究のための日本人 SFL 学習者による話し言葉コーパスの作成

ガルシア ルイス カステージョ・カルロス  
広島大学外国語教育研究センター

外国語学習におけるオーラル・インタラクションと日常会話の重要性にもかかわらず、日本人学習者が外国語としてのスペイン語（SFL）の会話に関する研究はまだありません。スペイン語と日本語の間に顕著な言語的距離があることを踏まえると、日本人 SFL 学習者の会話スタイルは、日本語やスペイン語を母語とする話者の会話スタイルとは異なると同時に、習熟度によっても変動すると言えます。

本稿では、会話分析と相互作用言語学の方法論に従って、日本人 SFL 学習者の会話スタイルの研究を行うことを目的として作成された、初の話し言葉コーパスを紹介します。まず、コーパスデータの設計と収集方針について議論します。次に、収集された会話から2つの事例を抜き出して紹介します。最後に、その2つの事例を分析します。この予備的分析の結果は、習熟度の低い参加者はより日本語に近いスタイルを使用し、ターン内のサポートターンまたは継承ターンの頻度が高いことを特徴とし、習熟度の高い参加者間の会話では、スペイン語での会話の特徴であるターンの終わりでのサポートターンの頻度の低さとオーバーラップの増加が特徴として見られることを示唆しています。