



# **The processing of gender information in languages of distinct morpho-syntactic gender marking systems**

Dissertation submitted to the  
Department of Basque Language and Communication  
of the  
University of the Basque Country  
in Candidacy for the  
Degree of  
Doctor of Linguistics

by  
Jui-Ju Su

Dissertation Director: Dr. Manuel Carreiras  
Co-Director: Dr. Nicola Molinaro

© 2017 by Jui-Ju Su  
All Rights Reserved.



*Dedicated to my beloved family and daughter.*

## Acknowledgement

Pursuing PhD is the most adventurous decision I've ever done in my life up to now as I had to go abroad with my daughter who was eight years old at that time. I've experienced a different life along the way. Thanks to the Lord, for his grace I have come to the very last step after nearly seven years.

First of all, I would like to thank Manuel Carreiras for giving me this opportunity to join the European Community's Seventh Framework Program (FP7/2007-2013), Marie Curie Initial Training Network – Language, Cognition and Gender (under grant agreement no. 237907). Without him, anything related to this dissertation would never happen. Thanks to his guidance on my academic study. He has led my career to a different level. A very special gratitude to Nicola Molinaro, my adviser. Thanks to his patience on my slow progress. Without him, this dissertation would not have accomplished. Thanks to all the professors, fellows and the generous funding in this Marie Curie ITN-LCG project.

I am thankful to Magaret Gillon-Dowens, Simona Mancini, Brendan Costello and Ainhoa Eguiguren for their unconditional help and constructive opinions on linguistic issues. Thanks to Lorna García Pentón for her invaluable friendship from the first day I arrived at the BCBL till now. Thanks to Daphne Li, Ya-Fen Chen, Kristi Tang, Huei-Chin Wu, Sharon Lin, Jessie Chi, Yi-Tien Tsai and Nor Chen for being my unfailing friends and giving me advices whenever I need. Thanks to Eider Juaristi, Ana Fernández, Vanessa Gallardo and Leire Arietaleanizbeascoa for their friendship and help during the time. Thanks to Aina Casaponsa, Oihana Vadillo, Larraitz López, Eri Takahashi, Borja Chantre and Xabier Rojo for their kind support on experimental and technical issues. Thanks to Joy Chien, Pei-Shu Tsai and Denise Wu for their considerate support during my stay at the National Central University, Taiwan. Thanks to all people at the BCBL. Thanks to everyone who participated in the experiments.

I am grateful to my family for their unconditional support so that I can stick to the end. Thanks to my beloved daughter for her accompany all the way. She is the reason that makes me move on. Thanks to the Lord for He has made all these possible.



# Abstract

Gender information is the basic feature about people which can be either explicitly expressed by definitional terms (e.g., father, mother, king or queen) or implicitly perceived from stereotypical role nouns (e.g., doctor, nurse, plumber or florist). Although both definitional and stereotypical role nouns can represent the same person, the two types of gender may sometimes collide and result in processing difficulties. Except for the two (explicitness of gender information: explicit vs. implicit), a third type of gender information is restrictedly required in some languages to linguistically represent the gender of people. Based on the presence/absence of morpho-syntactic gender and gender agreement, world languages are categorized into three groups: grammatical gender languages, natural gender languages, and genderless languages. Anaphoric pronoun resolution structure is the linguistic scenario used the most to study activation of gender information and most of these studies are tested in grammatical gender and natural gender languages. Not too much is known about genderless languages, especial processing of stereotypical knowledge.

The present dissertation is aimed to investigate gender processing in languages of distinct gender marking systems employing ERP mismatch paradigm. In Experiment one, anaphoric pronoun resolution is investigated in a genderless language, i.e., Mandarin Chinese. To be specific, except for the absence of morpho-syntactic gender and gender agreement, gender specificity is not symmetrically represented in Chinese third person pronouns (i.e., 他/*he* represents every individual and 她/*she* refers to females only). Thus, in addition to antecedents' gender explicitness (i.e., definitional and stereotypical), pronouns' gender specificity, is also manipulated to evaluate gender processing during Chinese anaphoric pronoun resolution.

In view of anaphoric pronoun resolution that measures brain activation on pronouns that are somehow few words after antecedents are read, a gap might exist if the observation is targeted on the role nouns. To understand better the immediate activation of stereotypical knowledge, the focus shifts from pronouns to online activation of stereotypical role nouns. Brain activation during the reading of stereotypical role nouns in sentences are studied in a grammatical gender language (e.g., Spanish, Experiment Two) and a genderless language (e.g., Chinese, Experiment Three).

In the last part of this dissertation, cross language comparisons are carried out respectively for anaphoric pronoun resolution and stereotypical role nouns. A mapping (Table 8) of the ERP components and the two-stage model for anaphoric pronoun resolution is done to fully illustrate this theoretical correlation in languages of differential gender markings. Regarding stereotypical knowledge, while it leads the processing in Spanish noun phrase, it can be modulated to some extent by gender markers in Chinese. The activation of stereotypical gender is lexical-driven in Spanish, but that in Chinese, is inferential and context dependent.

Overall, the present dissertation has shown different mechanisms related to gender processing in languages of distinct gender marking systems. Except for gender markings encoded in the testing languages, linguistic scenario (anaphoric pronoun resolution or role nouns) and target words (antecedents or pronouns) employed are also critical for such investigation.

## Resumen

La información de género es una característica básica sobre las personas que se puede expresar de forma explícita mediante términos definicionales (p. ej., padre, madre, rey o reina) o inferir de los sustantivos estereotípicos implícitos (p. ej., médico, enfermera, fontanero o florista). Tanto los sustantivos definicionales como los estereotípicos pueden representar a la misma persona. No obstante, a veces los dos tipos de géneros pueden chocar y dificultar el procesamiento, tal y como se muestra en el siguiente ejemplo:

*A car accident happened to a boy and his father. The father died. When the boy was sent to the hospital, the surgeon in the hospital saw the boy and said "Oh! My god, this is my son."*

(ejemplo extraído de Garnham y Oakhill, 2006, p. 886)

A la hora de leer este texto, los lectores tuvieron dificultades para interpretar quién (es decir, la madre) era *the surgeon* (sesgo masculino). Por tanto, la información de género puede estar sesgada, a nivel semántico, en la posición de antecedente y expresada de forma explícita mediante un sustantivo de rol definicional y/o de forma implícita mediante un sustantivo estereotípico. Además de estas dos maneras de expresar el género (explícita frente a implícita), en algunas lenguas existe un tercer tipo de información de género que representa el género de las personas a nivel lingüístico de forma obligatoria, es decir, el género morfosintáctico (masculino, femenino y/o neutro). Asimismo, requiere concordancia de género entre el género biológico de una persona y las categorías lingüísticas relacionadas que pudieran representar el género de la persona en cuestión (p. ej., sustantivos, pronombres, determinantes, adjetivos y, en algunas lenguas, verbos). En base a la ausencia/presencia de género morfosintáctico y concordancia de género, las lenguas del mundo se clasifican en tres grupos: lenguas de género gramatical, lenguas de género natural y lenguas sin género.

Por norma general, las lenguas de género gramatical incluyen «*las familias de lenguas eslavas (p. ej., ruso), germánicas (p. ej., alemán), romances (p. ej., español), indoarias (p. ej., hindi) o semíticas (p. ej., hebreo), entre otras*» (Stahlberg et al., 2007, p.164). En esta clase de lenguas, las características de género (masculino, femenino y/o neutro) se presentan mediante morfemas de género unidos a los sustantivos (tanto animados como inanimados) y la concordancia de género es necesaria en las palabras asociadas con un sustantivo concreto. Las lenguas de género natural no emplean marcadores de género gramaticales para los sustantivos, los cuales son casi genéricos y sirven para referirse a ambos géneros. La característica específica de una lengua de género natural es el pronombre de tercera persona (él o ella), que debería concordar morfosintácticamente con el género biológico de la persona en cuestión. La lengua inglesa y las escandinavas se clasifican como lenguas de género natural. En cuanto a las lenguas sin género, tal y como su nombre indica, ni los sustantivos ni los pronombres de tercera persona proporcionan información precisa sobre el género, ya que no requieren morfemas de género morfosintácticos o concordancia de género; las lenguas urálicas (p. ej., finés), turcas (p. ej., turco),

iraníes (p. ej., persa), siníticas (p. ej., chino) y bantúes (p. ej., suajili) pertenecen a este grupo de lenguas.

Dado que estamos especialmente interesados en cómo se procesa la información de género en el cerebro, la revisión bibliográfica se centra en los estudios mediante Potenciales Relacionados con Eventos o ERP (por sus siglas en inglés). Entre los estudios previos con ERP, el mecanismo lingüístico más empleado es la resolución de la anáfora pronominal y la mayoría se llevaron a cabo con lenguas de género gramatical y natural. No existe mucha información sobre las lenguas sin género, en especial sobre el procesamiento del conocimiento estereotípico. Estos estudios anteriores se centraron en cómo la información de género codificada en los antecedentes (de género gramatical definicional, estereotípico y/o morfosintáctico) influye en la resolución pronominal, centrándose en el procesamiento del pronombre o anáfora. A partir de los resultados, los investigadores sacaron conclusiones divergentes sobre la activación del género del antecedente. Mientras algunos estudios afirmaban que las reglas morfosintácticas de las lenguas limitan la activación de la información de género, otros estudios señalaron que la resolución de la anáfora está básicamente determinada por la sintaxis pero puede interactuar con la información del género semántico codificado en los antecedentes (definicionales o estereotípicos) durante la comprensión. Sin embargo, los resultados con ERP obtenidos de estos estudios también difirieron en cuanto a los grupos de lenguas. En inglés se registraron los efectos P600 mientras que, en el caso del alemán, se observaron los efectos N400 y P600. Al tener en cuenta la información contextual durante la resolución del pronombre (p. ej., el sesgo contextual ambiguo frente al no ambiguo hacia uno de los dos antecedentes), los resultados obtenidos de un estudio llevado a cabo sobre el holandés respaldan que la información referencial ambigua podría ocasionar una negatividad frontal continuada (Nref) asociada con los pronombres, al contrario que los no ambiguos.

Dos estudios recientes han examinado con ERP el procesamiento del género durante la resolución de la anáfora pronominal en chino, una lengua sin género. Sin embargo, debido a los objetivos propios de cada investigación, los resultados de ERP son ligeramente inconsistentes entre los dos estudios. Uno de los estudios (Qiu, Swaab, Chen y Wang, 2012) concluyó que el procesamiento de la información de género durante la resolución de la anáfora pronominal se «basa en la semántica» cuando el pronombre está más próximo al antecedente (es decir, N400 para distancias cortas) y que, además, esta representación disminuye a medida que aumenta la distancia (es decir, P600 para distancias largas). El otro estudio (Xu, Jiang y Zhou, 2013) registró solo efectos P600 y los interpretó como un reflejo de i) dificultad de integración general como resultado de los conflictos de valor del género durante la resolución de la anáfora en chino o ii) procesamiento «basado en la semántica» (es decir, computación de la relación semántica entre el antecedente y la anáfora) debido a la falta de concordancia de género morfosintáctico en chino. Sin embargo, solo se utilizaron nombres propios y sustantivos de roles de género específico en los estudios chinos, por lo que no se abordó el género estereotípico. Además, cabe destacar que no se tuvo en consideración la especificidad desigual de género de los dos pronombres en tercera persona del chino (es decir., 他/él se refiere a cualquier persona y 她/ella solo a la mujer).

En cuanto a la resolución de la anáfora, Garrod y Sanford (1994) y Garrod y Terras (2000) han propuesto un modelo en dos etapas en el que afirman que el procesamiento anafórico conlleva i) una etapa de vinculación y ii) una etapa de resolución durante la comprensión referencial. La etapa de vinculación ocurre en



primer lugar para filtrar el referente que mejor encaje de entre los posibles candidatos, es decir los antecedentes, en base a la información a nivel discursivo que incluye el género, el número y las reglas sintácticas. Una etapa de resolución posterior logra la interpretación de esa anáfora, teniendo en cuenta el conocimiento del mundo y la información contextual. Algunos investigadores han intentado incorporar la inferencia teórica del modelo en dos etapas a los descubrimientos obtenidos con ERP para explicar el mecanismo del procesamiento de género durante la resolución de la anáfora pronominal. Sin embargo, debido a la disparidad de los objetivos de investigación, el diseño de las investigaciones, los materiales de los experimentos, etc., la correlación teórica entre el modelo en dos etapas y los hallazgos con ERP se suele hacer en base a un solo estudio. Aprovechamos esta situación para contemplar la posibilidad de integrar estudios similares relacionados con el procesamiento de género para observar cómo varía esta correlación teórica según los distintos sistemas para marcar el género.

Por consiguiente, el objetivo de la presente tesis era investigar el procesamiento de género en lenguas con distintos sistemas para marcar el género mediante el paradigma de incongruencia con ERP. Respecto a la resolución de la anáfora pronominal, se han estudiado los sustantivos definicionales y estereotípicos en el caso de las lenguas de género gramatical (p. ej., alemán) y género natural (p. ej., inglés); aún no se han abordado los sustantivos estereotípicos de las lenguas sin género. En el primer experimento (Capítulo 2), se investiga la resolución de la anáfora pronominal en una lengua sin género, el chino mandarín. Más concretamente, además de la ausencia de género morfosintáctico y concordancia de género, la especificidad de género no está representada simétricamente en el caso de los pronombres de tercera persona del chino (p. ej., 他/*él* representa a cualquier persona y 她/*ella* corresponde solo a las mujeres). Por lo tanto, además de la explicitud del género de los antecedentes (es decir, definicional y estereotípico), la especificidad del género de los pronombres también se manipula para evaluar el procesamiento de género durante la resolución de la anáfora pronominal en el caso del chino. Para evitar faltas de comprensión con los pronombres simples de tercera persona, en este experimento se han utilizado pronombres reflexivos de tercera persona (p. ej., 他自己/*a sí mismo* vs. 她自己/*a sí misma*). Los resultados demostraron lo siguiente: en primer lugar, que la explicitud de género codificada en la posición antecedente no desempeña ningún papel durante la resolución de la anáfora pronominal; en segundo lugar, que el pronombre reflexivo 她自己/*a sí misma* consume más recursos cognitivos en un intervalo de tiempo perceptual temprano (P200) y un intervalo de tiempo semántico (N400), lo que respalda la suposición de que 他自己/*él mismo* sea el valor por defecto; en tercer lugar, surgió una disociación de la congruencia de género en el caso de 他自己/*a sí mismo* (P200) y 她自己/*a sí misma* (P600). Tal disociación aporta evidencias sobre la asimetría de la especificidad de género de la resolución de la anáfora y sobre los distintos cursos temporales involucrados en el procesamiento de género durante la resolución de la anáfora pronominal reflexiva.

En vista de la resolución del pronombre anafórico que mide la activación de los pronombres que aparecen algunas palabras después de los antecedentes, puede haber diferencias incluso cuando los participantes leen sustantivos de rol específico en la posición de antecedente; este fenómeno podría diferir según los distintos sistemas para marcar el género de las lenguas. En Carreiras, Garnham, Oakhill y Cain (1996), los autores utilizaron la resolución del pronombre anafórico para estudiar el procesamiento de género en sustantivos de rol específico en inglés. Los participantes

mostraron mayores tiempos de lectura a ritmo propio en los casos de incongruencias de género en la posición del pronombre en inglés y la posición del antecedente en español, respectivamente. De acuerdo con los autores, un género estereotípico «asumido» por defecto podría activarse inmediatamente después de que los participantes se encontraran con un sustantivo de rol, incluso sin ningún tipo de información de origen lingüístico (p. ej., se observan tiempos de lectura mayores para *la carpintera* frente a *el carpintero*, de sesgo masculino). Para entender mejor el procesamiento de género inmediato del conocimiento estereotípico, la atención cambia de los pronombres a la activación en línea de los sustantivos estereotípicos. Se ha investigado la activación cerebral durante la lectura de sustantivos estereotípicos en el caso de una lengua de género gramatical (p. ej., español, experimento dos) y una lengua sin género (p. ej., chino, experimento tres) para ver en qué difiere dicho procesamiento según el sistema para marcar el género.

En el experimento dos (Capítulo 3), aprovechamos las ventajas de la rica morfología de género gramatical del español para explorar el procesamiento de género de los sustantivos estereotípicos en las lenguas de género gramatical. Los estímulos estereotípicos y gramaticales se manipularon simultáneamente dentro de la misma frase nominal en un diseño factorial (es decir, determinante + sustantivo de rol, p. ej., los mineros/\*las mineros/\*los mineras/las mineras). De forma sorprendente, los resultados solo muestran efectos N400 en los sustantivos con incongruencia estereotípica y/o gramatical. Esto podría implicar que los morfemas de género morfosintáctico (es decir, -o/-a) presentes en los sustantivos estereotípicos de las lenguas de género gramatical funcionan meramente como indicio formal (gramatical y/o estereotípica) para que los participantes accedan al género estereotípico predeterminado. Por ello, en una lengua de género gramatical como el español, en la que el género gramatical es obligatorio, el conocimiento estereotípico anula las pistas sintácticas y destaca la inmediatez con la que se activa el conocimiento estereotípico durante la comprensión del lenguaje. Por tanto, respalda las propuestas que afirman que el conocimiento social impacta en el procesamiento del lenguaje de manera distinta que otras formas de semántica.

En el experimento tres (Capítulo 4), se investiga sobre el procesamiento de género de los sustantivos estereotípicos del chino mandarín, en el que no existen los morfemas de género morfosintáctico ni concordancia de género. El género de los sustantivos de rol del chino se expresan mediante una categoría encubierta de un sustantivo genérico (p. ej., 礦工/minero, masculino) y el marcador de género que precede al sustantivo de rol (p. ej., «男»礦工/minero «*masculino*»). El género estereotípico por defecto de un sustantivo de rol se percibe mediante la categoría encubierta, en lugar de basarse en ningún marcador de género formal. Es decir, cada sustantivo de rol genérico tiene un género por defecto por el cual se reconoce. Los marcadores de género son palabras independientes (p. ej., 男/masculino o 女/femenino) que deben aparecer antes (es decir, a la izquierda) de un sustantivo de rol genérico, si el género es contrario al predeterminado (p. ej., «女»礦工/minera «femenina»). No hay cambios en el sustantivo de rol genérico en sí y no se requiere concordancia de género morfosintáctico. Con el fin de estudiar cómo influye el sistema para marcar el género del chino en la comprensión de las frases nominales estereotípicas, se manipulan tres condiciones: predeterminado (p. ej., 礦工/minero, sesgo masculino), estereotipo congruente (p. ej., 男的礦工/minero *masculino*) y estereotipo incongruente (p. ej., 女的礦工/minera *femenina*). Los resultados indicaron un P200 relacionado con la atención y un efecto P600 relacionado con la

integración para los predeterminados, al contrario que las otras dos condiciones de género específico, y un efecto del tipo N400 caracterizó la condición incongruente comparada con la predeterminada. Además, se produjo un efecto muy tardío (probablemente procesamiento «post-semántico», 800–1000ms) en el caso de las frases nominales estereotípicamente incongruentes, en comparación con las predeterminadas y congruentes. Los resultados sugieren que la inmediatez del conocimiento estereotípico es similar en todos los idiomas. Sin embargo, el uso de una categoría encubierta, junto con la presencia/ausencia de marcadores de género, puede modular el efecto inmediato del conocimiento estereotípico en el caso del idioma chino. Por tanto, la activación del conocimiento estereotípico en chino es inferencial y dependiente del contexto, pero no está determinado a nivel léxico por la raíz de la palabra, tal y como se observa en las lenguas con morfología rica.

En la última parte de esta tesis (Capítulo 5), se llevan a cabo comparaciones entre lenguas, tanto de la resolución de la anáfora pronominal como de los sustantivos de rol específico. Asimismo, se ha realizado un mapeo del modelo en dos etapas de la resolución del pronombre anafórico y los hallazgos con ERP para ilustrar en profundidad esta correlación teórica en las lenguas con distintas maneras de marcar el género. De acuerdo con la Tabla 8, el componente P200 está relacionado con la etapa de vinculación, mientras que los componentes N400 y P600 están estrechamente ligados a la etapa de resolución. Sin embargo, la actividad durante la etapa de resolución podría diferir según los distintos sistemas para marcar el género y, además, se podrían distinguir dos fases subordinadas: una fase inicial sujeta a posibles interferencias y una fase de integración que cierra el proceso. En la «fase de interferencia» inicial, la interferencia puede estar determinada por la información codificada en el antecedente (observado en el caso del alemán), o en el mismo pronombre (como en el caso del chino). La fase final de integración es el último paso en la etapa de resolución, que representa la resolución de cualquier información relacionada. Los efectos P600 en el caso de los pronombres incongruentes de género observados en algunos estudios previos son prueba de esta fase. Por otro lado, la activación inmediata de conocimiento estereotípico se ha observado en lenguas con sistemas para marcar el género distintos. Sin embargo, mientras que el conocimiento estereotípico domina en el procesamiento de las frases nominales en español, se puede modular hasta cierto punto mediante marcadores de género en el caso del chino. La activación del género estereotípico está lexicalmente determinada en español mientras que, en chino, es inferencial y depende del contexto.

En general, la presente tesis ha mostrado mecanismos diferentes relacionados con el procesamiento del género en lenguas con sistemas para marcar el género distintos. Además de la forma de codificar el género en las lenguas testadas, el marco lingüístico (resolución del pronombre anafórico o sustantivos de rol) y las palabras-objetivo (antecedentes o pronombres) que se han utilizado son cruciales para este tipo de investigación.



# Table of Contents

Acknowledgement .....	ii
Abstract .....	iv
Resumen.....	v
Table of Contents .....	xi
List of Tables .....	xiv
List of Figures .....	xiv
CHAPTER 1 .....	1
Introduction of the Present Dissertation .....	1
1.1 Gender in Languages .....	2
1.1.1 Definitional Gender vs. Stereotypical Gender .....	2
1.1.2 The Linguistic Coding of Gender Information in Languages.....	3
1.2 An Overview of Previous Studies on the Processing of Gender Information .	4
1.2.1 Studies on Non-animate Gender Processing.....	4
1.2.2 Studies on Animate Gender Processing.....	5
1.2.3 Some Aspects that Have not Been Addressed in Previous Studies .....	7
1.3 The Present Dissertation .....	11
1.3.1 Experiment I: Anaphoric Pronoun Resolution in A Language Where Gender is not Morpho-syntactically Encoded.....	11
1.3.2 Experiment II: The Online Processing of Stereotypical Gender in Grammatical Gender Language .....	12
1.3.3 Experiment III: The Online Processing of Stereotypical Gender in Genderless Language .....	12
1.3.4 Summary of Chapter 1 .....	13
CHAPTER 2 .....	14
Experiment I: Pronoun Resolution in A Language Where Gender Is not Morpho- syntactically Encoded .....	14
2.1 Introduction.....	15
2.2 Materials and Methods.....	21
2.2.1 Participants.....	21
2.2.2 Materials .....	21
2.2.3 Procedure .....	23
2.2.4 Data Recoding and Analysis.....	24
2.3 Results.....	24
2.3.1 ERPs on the Reflexive Pronouns at the P200 Time Interval: 150 ms to 250 ms.....	25
2.3.2 ERPs on the Reflexive Pronouns at the N400 Time Interval: 250 ms to 600 ms.....	27
2.3.3 ERPs on the Reflexive Pronouns at the P600 Time Interval: 600 ms to 800 ms.....	28
2.4 Discussion .....	29
2.4.1 The Processing of Chinese Third Person Reflexive Pronouns .....	30
2.4.2 Anaphor Resolution in a Language without Inflectional Morphological Gender Markings .....	31
2.5 Conclusion .....	33
2.6 Problems Unsolved and the Following Experiment .....	34
CHAPTER 3 .....	36

Experiment II: The Online Processing of Stereotypical Gender in a Grammatical Gender Language .....	36
3.1 Introduction.....	37
3.2 Materials and Methods.....	40
3.2.1 Participants.....	40
3.2.2 Materials .....	40
3.2.3 Procedure .....	40
3.2.4 Data Recording and Analysis.....	40
3.3 Results.....	41
3.3.1 Comprehension Questions .....	41
3.3.2 ERPs Time-locked to Role-nouns.....	41
3.3.2.1 The 300-600 ms Interval.....	41
3.3.2.2 The 600-850 ms Interval.....	42
3.4 Discussion .....	44
3.5 Problems Unsolved and the Following Experiment .....	45
CHAPTER 4 .....	47
Experiment III: The Online Processing of Stereotypical Gender in a Genderless Language.....	47
4.1 Introduction.....	48
4.2 Methods.....	52
4.2.1 Participants.....	52
4.2.2 Materials .....	52
4.2.3 Procedure .....	53
4.2.4 Data Recording and Analysis.....	53
4.3 Results.....	54
4.3.1 Comprehension Questions .....	54
4.3.2 ERPs on the Role nouns.....	54
4.3.2.1 The P200 Time Interval (130–250 ms).....	57
4.3.2.2 The N400 Time Interval (250–600 ms) .....	57
4.3.2.3 The P600 Time Interval (600–800 ms and 800–1000 ms) ...	57
4.4 Discussion .....	60
4.4.1 Processing Modulation Given by Gender Marking: The P200.....	60
4.4.2 Processing Modulation Given by Gender Marking: The P600.....	61
4.4.3 Stereotypical congruence: The case of Chinese.....	62
4.5 Conclusion .....	63
CHAPTER 5 .....	64
Overall Discussion of the Present Dissertation.....	64
5.1 Summary of the Results in Each Experiment .....	65
5.2 The Processing of Gender Information during Anaphoric Pronoun Resolution across Languages with Distinct Gender Marking Systems.....	66
5.2.1 The ERP Correlates Reported in Studies Involving Anaphor Resolution across Languages with Distinct Gender Marking Systems .....	66
5.2.2 The ERP Correlates and the Two-stage Model for Anaphoric Pronoun Resolution across Languages with Distinct Gender Marking Systems .....	68
5.3 The Processing of Stereotypical Gender in Noun Phrases across Different Gender Systems .....	70
5.4 Future Directions on the Processing of Gender Information.....	72
5.5 Conclusion .....	75
References.....	76
Appendix A. Proper names (definitional role nouns) used in Experiment I .....	84

Appendix B.	Stereotypical role nouns used in Experiment I .....	85
Appendix C.	Male-biased rating scores for the 80 Spanish role nouns employed in the present experiment. ....	86

## List of Tables

Table 1. Word Frequencies for the Two Chinese Pronouns Extracted from Word List with Cumulated Word Frequency in Sinica Corpus, Academia Sinica, Taiwan. ...	18
Table 2. Example Sentences Used in the Experiment .....	23
Table 3. Example of the Experimental Sentences .....	39
Table 4. The design reported in Molinaro, Su and Carreiras (2016): the lower part in each condition is the result when the condition is compared with the control condition.....	49
Table 5. Experimental conditions and example sentences presented in the experiment.....	51
Table 6. Mean amplitudes and <i>F</i> values resulting from Post hoc pairwise comparisons between each two manipulated conditions of STEREOTYPICAL GENDER CONGRUITY in the specific topographic regions respectively in the P200, N400 and P600 time intervals.....	59
Table 7. List of ERP Studies on the Processing of Gender Information during Anaphoric Pronoun Resolution.....	67
Table 8. The inferred mapping of ERP Components and the two-stage model in each language group.....	70
Table 9. The summary of the ERP components found in the present dissertation and some previous representative studies related to the processing of gender information.....	74

## List of Figures

Figure 1. Overall averaged brain activities of reflexive pronoun gender specificity (general vs. specific) by gender congruity (congruent vs. incongruent) in the representative electrodes. ....	25
Figure 2. The averaged brain activities separately presented by reflexive pronoun gender specificity (default: 他自己/ <i>himself</i> vs. specific: 她自己/ <i>herself</i> ) in the representative electrodes. ....	26
Figure 3. Topographic distributions for the reflexive pronoun gender specificity effects (她自己/ <i>herself</i> minus 他自己/ <i>himself</i> ) respectively in the P200 and N400 time intervals.....	27
Figure 4. Topographic distribution of the gender congruity effect (incongruent minus congruent) respectively for the two reflexive pronouns in the P200 and P600 time intervals. Panel (A) represents effects for the general reflexive pronoun, 他自己/ <i>himself</i> , and panel (B) represents effects for the specific reflexive pronoun, 她自己/ <i>herself</i> . ....	29
Figure 5. ERPs time-locked to the presentation of the target stereotypical noun. In the upper panel the four conditions are plotted in the same graph for two relevant electrodes. In the intermediate panels the topographical scalp distributions of each anomalous condition compared to the control in the 300–600 ms and 600–850 ms time intervals are plotted. In the lower panel we highlight the relevant effects of interest. ....	43
Figure 6. 1A: ERPs time-locked to the target words of the three experimental conditions of Stereotypical Gender Congruity at nine representative electrodes in the frontal, central and posterior scalp regions. 1B: Representative electrode (Fz) in the mid-anterior region respectively for male biased role nouns and female	



biased role nouns.....	55
Figure 7. Topographic distribution of the pair-wised congruity effect (incongruent minus congruent or congruent/incongruent minus default) respectively in the P200 (2A), N400 (2B) and P600 (2C) time intervals.....	56

# **CHAPTER 1**

## **Introduction of the Present Dissertation**

## **1.1 Gender in Languages**

When seeing a newborn baby, people always ask, is the baby a boy/he or a girl/she? How old is this baby? Or who does he/she look like, the father or the mother? ... etc. Among all these questions, the answer to the first question is a never changing fact in the baby's life. The answer is either male or female and cannot be both at the same time. It means that gender is the basic, hardly changing feature of a person. This might be the reason why gender information is consciously or unconsciously (maybe habitually) addressed in daily life when talking about the biological nature of people.

On the other hand, when talking about people's identity, various identities could represent a person at the same time. For instance, a man can be a husband, a father, an uncle, an architect, a fan of automobile race, ... and an after-hour columnist. In this case, gender information is more or less explicitly or implicitly expressed by these identities. Here, the consciousness and explicitness towards gender is demonstrated by referring to a person's biological gender, (i.e., definitional role noun; e.g. husband or father), and occupational identity (i.e., stereotypical role noun, architect or a fan or automobile race) of this person. Both the definitional and stereotypical role nouns can provide information on gender and the present dissertation is focused on how such gender information is processed in languages with distinct gender markings systems.

### **1.1.1 Definitional Gender vs. Stereotypical Gender**

Definitional gender and stereotypical gender are two types of gender information that can be used to refer to the same person. However, there remain some subtle differences between the two and this could result in big differences when applied to language (i.e., the linguistic context of anaphoric/cataphoric resolution, see Duffy & Keir, 2004; Kreiner, Sturt & Garrod, 2008).

As the term implies, definitional gender refers to the biological gender, that needs to be learned for each role noun. For instance, king, father, uncle, or brother is male, and queen, mother, aunt, or sister is female. At the time when a role noun is used to represent a person, its gender is clearly specified as either male or female.

Stereotypical gender, instead, is an acquired conceptual knowledge related to a certain role noun in society (Martin & Rubel, 2004). Although a general role noun can be used for both genders, stereotypical gender is the biased gender that can represent the majority of this role in society. For example, a miner can be a male or a female, but in general, miners are male. Stereotypical gender of a certain role noun thus is not defined but framed by society. Different from definitional gender that is learned and consistent across languages, societies and time, stereotypical gender, on the other hand, is acquired and perceived by people from society: it may be different across language communities and it can change along with the evolution of a society. For instance, medical degree and law earners were about 10 percent females in the 1970's and the percentage had reached up to nearly 50 percent in 2010 in the United States (Cohen, 2012). In the field of medicine, women became the majority in the pediatrics and obstetrics departments in 2015 in the United States (Grover, 2015).

Although both definitional and stereotypical role nouns can represent the same person, the two types of gender may sometimes collide and result in processing difficulties as shown in (1). When reading this paragraph, readers encountered difficulties while interpreting who the surgeon (male-biased) was (i.e., the mother). Gender information thus can be semantically interpreted or biased in the antecedent

position explicitly expressed by a definitional role noun and/or implicitly by a stereotypical role noun.

- (1) *A car accident happened to a boy and his father. The father died. When the boy was sent to the hospital, the surgeon in the hospital saw the boy and said “Oh! My god, this is my son.”*

(example taken from Reynolds, Garnham, & Oakhill, 2006, p. 886)

Except for biological gender (both definitional and stereotypical) that could represent the sex of a person, linguistic forms (i.e., the so-called grammatical gender) can also indicate the sex of a person at the same time. But, such linguistic forms may also intervene and make the processing of gender more complicated when there are conflicts among them (e.g., see Irmen & Roßberg, 2004 and Irmen, Holt & Weisbrod, 2010, for the processing of biological gender, stereotypical gender and grammatical gender during anaphoric resolution in German).

### 1.1.2 The Linguistic Coding of Gender Information in Languages

The linguistic implementation of gender information varies across languages. Based on their grammatical features, languages are categorized into three groups: grammatical gender languages, natural gender languages and genderless languages (Stahlberg, Braun, Irmen & Sczesny, 2007).

In general, grammatical gender languages include “*the language families such as Slavic (e.g., Russian), Germanic (e.g., German), Romance (e.g., Spanish), Indo-Aryan (e.g., Hindi), or Semitic (e.g., Hebrew) and others*” (Stahlberg et al., 2007, p.164). Gender features (masculine, feminine, and/or neuter) in grammatical languages are presented via gender morphemes bound to the nouns (both animate or inanimate), and gender agreement is required for words related to a certain noun (e.g., determiners, adjectives, pronouns and/or verbs ... etc., Corbett, 1991).

Natural gender languages have no grammatical gender markings on the nouns and the nouns are almost generic, referring to both genders. The specific feature that includes a language in this category is the third person pronoun (he or she) that should morpho-syntactically agree with the biological gender of its antecedent. English and Scandinavian languages are categorized in this language group (Stahlberg, Braun, Irmen & Sczesny, 2007).

In genderless languages, as indicated by the term, neither nouns, nor third person pronouns provide clear information about gender as no morpho-syntactic gender morphemes are part of these words. Especially, in Mandarin Chinese for instance, the pronunciation for 他/*he* and 她/*she* are identical as /*ta*/. It happens very often in daily life, when the gender of a person is not clear from an utterance and the ambiguity can last for a while till the gender of the person is clarified, as shown in the Chinese example (2). Even so, gender information still can be expressed via semantic meanings, such as adjectives, 男的/*nán de* (*male*) or 女的/*nǚ de* (*female*) in Chinese. “*Language families such as Uralic (e.g., Finnish), Turkic (e.g., Turkish), Iranian (e.g., Persian), Sinitic (e.g., Chinese), and Bantu (e.g., Swahili)*” belong to this language group (Stahlberg, Braun, Irmen & Sczesny, 2007, p.166).

- (2) A: 今天班上來了一位新同學，/ta/跟我一起搭校車，/ta/似乎是個很不錯的人，我想接下來的日子我應該會跟/tā/相處愉快。

B: /ta/是男的還是女的？

A: 女的。

[A: A new classmate came to our class today. /ta/ (He/She) was sitting next to me at school bus and /ta/ (he/she) seems to be a nice person. I think I will get along well with /ta/ (him/her) in the future.

B: Is /ta/ male or female?

A: Female.]

## **1.2 An Overview of Previous Studies on the Processing of Gender Information**

Regarding the processing of gender information, abundant studies can be listed in the literature. In general, the focus of these studies can be divided into two categories, one is related to the processing of non-animate nouns (i.e., object nouns) and the other is focused on animate nouns (i.e., mainly role nouns referring to people). The two types of studies are discussed in the following sections with a specific emphasis on ERP (Event Related Potentials) studies.

ERPs are scalp recordings of electrical brain activity time-locked to a stimulus event. Compared to other neuroimaging techniques, they offer very good temporal details of the brain activity (in the order of tens of milliseconds). This makes the ERP technique one of the best measures for disentangling the temporally incremental neural processes typically assumed by cognitive models.

In the literature, three ERP components, i.e., the LAN (Left Anterior Negativity), the N400 and the P600 effects, are highly correlated with language processing. The N400 component is a negative peak observed around 400 ms post-stimulus onset distributed mostly in the central-to-posterior regions of the scalp. According to the literature, the N400 reflects context-dependent lexico-semantic processing of a given stimulus because it is modulated depending on a single word's lexical properties and the contextual information provided by the previous context (for a review, Kutas & Federmeier, 2011). The P600 component is a positive shift observed after 500 ms post-stimulus onset and its distribution is across the whole scalp. It correlates with general integration difficulties during sentence processing (either semantic or grammatical), and is typically observed in manipulations involving grammatical complexities, grammatical ambiguities or errors (Osterhout & Holcomb, 1992; Munte et al., 1998; Kaan & Swaab, 2003; Carreiras, Sallillas & Barber, 2004; Van Herten, Kolk & Chwilla, 2005; Kuperberg et al., 2006; Molinaro, Carreiras & Duñabeitia, 2012). In studies related to grammatical agreement, a P600 effect is often found to be preceded by a LAN peaking around 350 ms post-stimulus onset distributed in the left anterior scalp region when two words morpho-syntactically disagree (Barber & Carreiras, 2005; Molinaro et al., 2008b; 2011a, 2011b; 2015; O'Rourke & Van Petten, 2011; Regel, Meyer & Gunter, 2014; Caffara et al., 2015). The available findings related to the processing of grammatical gender errors indicate that the LAN is mainly reflecting early automatic syntactic processing, while the P600 is reflecting sentence-level integration difficulties.

### **1.2.1 Studies on Non-animate Gender Processing**

Most ERP studies on gender processing involving non-animate entities have been carried out in grammatical gender languages (Molinaro, Barber & Carreiras, 2011; Caffarra et al., 2015, and others). Grammatical gender in these studies was

employed as a cue to evaluate the grammaticality of a word (or a noun phrase). Overall, these studies usually investigated the relationship between semantic and syntactic processing, and the target words were always nouns or adjectives in which the gender value should morpho-syntactically agree with another word.

Some studies employing nouns referring to non-animate entities have provided evidence that semantic anomalies elicit N400 effects and grammatical gender violations elicit P600 effects on the target word (Hagoort and Brown, 1999; Wicha, Bates, Moreno & Kutas, 2003; Wicha, Moreno & Kutas, 2003). In addition, consistent LAN plus P600 effects have been reported in some studies when grammatical gender was mismatching within noun phrases (Gunter, Friederici & Schriefers, 2000; Barber & Carreiras, 2005; Molinaro, Vespignani & Job, 2008; O'Rourke & Van Petten, 2011; Caffarra, Siyanova-Chanturia, Pesciarelli, Vespignani & Cacciari, 2015).

### 1.2.2 Studies on Animate Gender Processing

Studies on gender processing referring to animate entities, on the other hand, focused on the processing of human role nouns (i.e., definitional and/or stereotypical). Among the studies, pronoun resolution is the linguistic structure used the most. These studies mainly evaluated how the gender information encoded in the antecedents (definitional, stereotypical and/or the morpho-syntactic grammatical gender) influences pronoun resolution and the focus was on the processing of pronoun or anaphora. Researchers made inferences about gender activation of antecedents from the results. Studies have been mostly carried out in grammatical gender languages (Dutch: Nieuwland & Van Berkum, 2006; German: Schmitt, Lamers & Münte, 2002; Irmen & Roßberg, 2004; Irmen, Holt & Weisbrod, 2010; Esaulova, Reali & von Stockhausen, 2014; Spanish and French: Garnham, Oakhill, Ehrlich & Carreiras, 1995; Spanish: Carreiras, Garnham & Oakhill, 1993; Carreiras, Garnham, Oakhill & Cain, 1996; Barber, Salillas & Carreiras, 2004; ...etc. and others) and natural gender languages (English: Osterhout & Mobley, 1995; Osterhout, Bersick & McLaughlin, 1997; Garnham, Oakhill & Reynolds, 2002; Kennison & Trofe, 2003; Sturt, 2003; Duffy & Keir, 2004; Oakhill, Garnham & Reynolds, 2005; Reynolds, Garnham & Oakhill, 2006; Kreiner, Sturt & Garrod, 2008; Kreiner, Mohr, Kessler & Garrod, 2009; ...etc. and others). Few have been done in genderless languages (Finnish: Pyykkönen, Hyönä & van Gompel, 2010; Chinese: Huang & Luh, 2012; Qiu, Swaab, Chen & Wang, 2012; Xu, Jiang & Zhou, 2013).

For the processing of human role nouns (either definitional or stereotypical or both) employing pronoun resolution, divergent results have been reported in previous studies. While most studies on anaphoric pronoun resolution have claimed that morpho-syntactic rules in languages constrain the activation of gender information (Spanish: Carreiras, Garnham & Oakhill, 1993; Spanish and English: Carreiras, Garnham, Oakhill & Cain, 1996; Spanish and French: Garnham, Oakhill, Ehrlich & Carreiras, 1995; English: Osterhout, Bersick & McLaughlin, 1997; Kennison & Trofe, 2003; Sturt, 2003; Duffy & Keir, 2004; Kreiner, Sturt & Garrod, 2008; Kreiner, Mohr, Kessler & Garrod, 2009; German: Irmen & Roßberg, 2004; Esaulova et al., 2014), some other studies have reported that anaphor resolution is basically syntactical-driven but can interact with semantic gender information (either definitional or stereotypical) during language comprehension (German: Schmitt, Lamers & Münte, 2002; Irmen, Holt & Weisbrod, 2010). For those studies employing ERP mismatch paradigm with antecedents of definitional and/or stereotypical role nouns, while the P600 effects are reported in English (natural gender language: Osterhout, Bersick & McLaughlin, 1997; Kreiner, Mohr, Kessler & Garrod, 2009), N400 plus P600 effects

are reported in German (grammatical gender language: Schmitt, Lamers & Münte, 2002; Irmen, Holt & Weisbrod, 2010). When contextual information was also taken into account (e.g., the ambiguous vs. unambiguous contextual bias towards one of the two antecedents) during pronoun resolution, the results obtained from a Dutch study have provided consistent evidence with previous studies that ambiguous referential information would cause sustained frontal negative shift (Nref) on the pronouns as opposed to the non-ambiguous pronouns (Nieuwland & Van Berkum, 2006).

In this respect, Garrod and Sanford (1994), and Garrod and Terras (2000) have proposed a two-stage model stating that anaphoric processing involves (i) a bonding stage and (ii) a resolution stage during referential comprehension. A bonding stage occurs first for the purpose to sort for a best-fit referent among possible candidates, i.e. the antecedents, based on the discourse level information, including gender, number and syntactic rules. A resolution stage subsequently accomplishes the interpretation of that anaphora taking into account world knowledge and contextual information.

Some studies have tried to correlate morpho-syntactic rules, semantic information, world knowledge and/or contextual information respectively with the bonding and the resolution stages. Sturt (2003) employed two eye-tracking experiments (in English, a natural gender language) and found that morpho-syntactic information is processed at the bonding stage (including stereotypical gender information), and the contextual information is processed later at the resolution stage. It should be noted here that the gender information is part of syntactic rules for English reference resolution (a natural gender language). However, in another eye-tracking study carried out by Esaulova, Reali and von Stockhausen (2014) in German (a grammatical gender language), the authors concluded that grammatical morpho-syntactic rules are processed at the bonding stage and the semantic feature (i.e., world knowledge, such as stereotypical gender) is processed later at the resolution stage. Thus, the processing of stereotypical gender differed in the two eye-tracking studies. While stereotypical gender is claimed to be processed at the bonding stage in English, in German it is referred to be processed at the resolution stage. Such difference might be attributed to distinct morpho-syntactic gender marking systems encoded in English and German.

ERP studies focused on anaphoric pronoun resolution have been carried out in genderless languages recently, (i.e., Chinese: Qiu, Swaab, Chen & Wang, 2012; Xu, Jiang & Zhou, 2013). In Qiu and colleagues (2012) two factors were manipulated: gender congruency and the distance between the antecedent and its pronoun. The results showed respectively N400 and P600 mismatch effects for short and long distance manipulations on the target pronouns. The authors concluded that the processing of gender information during Chinese anaphoric pronoun resolution is “semantics-based” when the pronoun is closer to the antecedent (i.e., N400 found for short distance) and this representation decays as the distance increases (i.e., P600 found for long distance). Similar long distance manipulation was employed in Xu and colleagues (2013, Experiment 1, singular antecedent): P600 effects were reported for gender mismatches on the target pronouns. The authors interpreted this P600 effect as reflecting either (i) general integration difficulty resulting from gender value conflicts during Chinese anaphor resolution or (ii) “semantics-based” processing (i.e., computation of the semantic relationship between antecedent and anaphora) due to the lack of morpho-syntactic gender agreement in Chinese. Importantly, only proper names and gender specified role nouns were used in these two Chinese studies, while other sources of gender-related information were not investigated (to be discussed at

the following section, 1.2.3) and the theoretical correlation between the ERP findings and the two-stage model are not clearly addressed.

A special case is the eye-tracking study by Pyykkönen, Hyönä and van Gompel (2010) in Finnish (a genderless language with only one pronoun for masculine and feminine, but with morpho-syntactic third person anaphoric verb that should agree with a preceded third person who could be either male or female). In this study, a visual world paradigm (participants stare at a monitor with different pictures while listening to speech stimuli) was used while monitoring participants' eye movements towards characters (a male or a female); meanwhile, they were hearing sentences containing stereotypical role nouns. According to the eye-tracking paradigm, the more fixation time the participants spend on a picture/character, the more information the participants get from the acoustic stimulus in relation to the picture. The results showed that structural salience (i.e., the order of mention) is reflected on the third person anaphoric verb; and a bias towards stereotypical gender was observed on stereotypical role nouns. The authors concluded that stereotypical gender is activated even if no formal linguistic gender agreement is required.

Apart from anaphoric pronoun resolution, gender information, especially stereotypical gender, has also been studied via the priming paradigm (English: Banaji, Hardin & Rothman, 1993; Banaji & Hardin, 1996; White, Crites, Taylor & Corral, 2009; Italian: Cacciari and Padovani, 2007; Siyanova-Chanturia, Pesciarelli & Cacciari, 2012; Chinese: Wang, Yang & Zhao, 2010) or manipulation of speakers' voices (Dutch: Van Berkum, van den Brink, Tesink, Kos & Hagoort, 2008; Slovakian: Hanulikova & Carreiras, 2015). The relationship between gender-related prime words (e.g., men, women, doctors or nurses) and target words (e.g., pronouns, reflexives, or gender trait-related adjectives such as *dependent* or *aggressive*) are manipulated in the priming studies. The results showed longer reaction times for gender mismatched target words than gender matched ones. It thus seems that the activation of stereotypical gender is automatic and immediate upon participants perceiving a role noun (Banaji, Hardin & Rothman, 1993; Banaji & Hardin, 1996). Interestingly, in those studies using ERP technique (with the mismatch paradigm), larger N400 effects were consistently reported for stereotypical gender mismatches on the target words that were primed by stereotypical role nouns (White, Crites, Taylor & Corral, 2009; Siyanova-Chanturia, Pesciarelli & Cacciari, 2012; Wang, Yang & Zhao, 2010).

In the second set of studies, speaker's voice (either male voice or female voice) was also manipulated to evaluate gender stereotyping for the message containing gender-biased information while participants listened to the messages. Robust semantic-related N400 effects were also found for stereotypical gender inconsistent messages in two ERP studies (Van Berkum, van den Brink, Tesink, Kos & Hagoort, 2008; Hanulikova & Carreiras, 2015).

### **1.2.3 Some Aspects that Have not Been Addressed in Previous Studies**

As mentioned in sections 1.2.1/2, most studies related to the processing of gender information focused on grammatical gender languages and natural gender languages. Not too much is known in genderless languages with poor or absent morpho-syntactic gender markings. Among previous studies, while consistent findings (N400 semantic mismatch effects) have been reported in studies employing the priming paradigm (at word level) and speakers' acoustic voices, results for studies involving anaphoric pronoun resolution are not consistent within/across language groups (i.e., grammatical gender languages, natural gender languages and genderless languages). After a closer look at these studies, the main focus of each study differed



(e.g., how gender information is processed, contextual bias, working memory, semantic or syntactic interference... etc.). Although the same syntactic construction (anaphoric pronoun resolution) was used, the number of antecedents, the order of mention, the manipulated referring features (e.g., syntactic or semantic), gender features of antecedents (i.e., role nouns) themselves (e.g., definitional, stereotypical or diminutive), and even the involvement of grammatical gender features (e.g., masculine, feminine or neuter) largely varied.

For the purpose of properly estimating gender processing during anaphor resolution across the three language groups of distinct gender marking systems, a set of studies with similar research question, sentence structure and experimental paradigm must be considered. Here, we are particularly interested in determining how gender information is resolved during anaphor resolution whose antecedent's gender is explicitly encoded in definitional role nouns and implicitly represented in stereotypical role nouns: first, we need to consider a set of studies investigating anaphor resolution with antecedents of either definitional or stereotypical or both, available in each of the three language groups. Second, we focus on studies with only one antecedent in each sentence to properly compare how antecedent's gender explicitness influences anaphor resolution. Third, we consider studies employing ERP mismatch paradigm to properly compare the mechanisms and time course of gender processing during anaphor resolution across the three language groups. Taken together, six studies corresponding to the three criteria mentioned above are suitable for estimating a cross-language comparisons. One study is however missing, i.e., an ERP study with antecedent of stereotypical role nouns in a genderless language.

Among these comparable ERP studies involving anaphoric pronoun resolution, the one carried out by Osterhout, Bersick and McLaughlin (1997) is a classic example, with one antecedent preceding its anaphora in a sentence. In this study, the authors manipulated antecedents with explicit gender (definitional role nouns) and implicit gender (stereotypical role nouns), and recorded participants' brain response to the pronouns in sentences. P600 effects were found in both types of antecedents when the gender of a pronoun mismatched the gender of its antecedent. Although the P600 effect found in sentences with definitional role nouns was larger than that found in sentences with stereotypical ones, given similar P600 topographic distributions, the authors concluded that the gender information was encoded in grammar independently of its explicitness in the role nouns.

Nonetheless, the authors did not provide a clear theoretical interpretation of how gender information is processed during pronoun resolution. For instance, whether the results support the two-stage model (bonding stage and resolution stage) proposed by Garrod and Sanford (1994) is not discussed.

On the other hand, in a genderless language (i.e, Mandarin Chinese) similar P600 effects are reported during anaphoric pronoun resolution with antecedents whose gender is definitional and in long-distance manipulations (Qiu, Swaab, Chen & Wang, 2012; Xu, Jiang & Zhou, 2013, Experiment 1). Here however, they are respectively interpreted as reflecting either working memory decay or semantics-based integration difficulty. For example, the P600 effect found in Xu and colleagues (2013) was interpreted as reflecting resolution stage. Such conclusion was reached based on Callahan (2008): the LAN (left anterior negativity) and the Nref (sustained anterior negativity) would reflect the bonding stage, while N400 and P600 would represent the resolution stage. Accordingly, all these P600 effects found for gender-mismatched pronouns may be interpreted as resulting from syntactic or semantic processing difficulties, thus reflecting more the resolution stage. Since no LAN or Nref have

been reported during anaphoric pronoun resolution with one antecedent only (showing no bonding difficulties), how gender information is processed at the bonding stage is still not clear. Specifically, we ask if the mapping between the two-stage model and the ERP components differ by the type of antecedent and/or by distinct gender marking systems of our interests.

In addition to these ERPs-theory mapping issues in all of these studies, pronouns typically are the target words on which relevant effects are measured and from which inferences about stereotype processing are made. Consequently, the gender specificity encoded in the pronouns should be crucial for such process. Gender specificity in pronouns in most Indo-European languages is clearly recognizable, i.e., “*he*” refers to male and “*she*” to female. Interestingly, the gender specificity in Chinese third person pronouns is not equally specified (他/*he* generally refers to every individual and 她/*she* is specifically used for female third person). This gender specificity feature has never been addressed in the literature and it might lead to different processing patterns during anaphoric pronoun resolution.

Furthermore, studies involving the processing of gender information (especially stereotypical gender information) focused more on anaphoric pronoun resolution in grammatical languages and natural languages. Related to this, Pyykkönen, Hyönä and van Gompel (2010) have pointed out that “*Investigating elaborative gender-stereotype inferences during discourse processing has been difficult [...], because they usually require a mismatch between a stereotype noun and another expression (e.g., a gender-marked pronoun) appearing later in the discourse. The activation of gender-stereotype information at this later point may be the result of a bridging inference rather than elaborative activation of stereotype information*” (p. 132). This is to say: pronoun resolution could only provide indirect evidence of stereotypical processing because the target pronoun is located several words after the antecedent where stereotypical gender information is activated. Since the activation of stereotypical gender is automatic and immediate (Banaji, Hardin & Rothman, 1993; Banaji & Hardin, 1996), it is crucial to evaluate the “immediate” online processing of stereotypical knowledge while participants are reading the role nouns.

More issues: throughout the literature, very few sentence processing studies directly looked at the processing of stereotypical gender time-locked to the comprehension of the role nouns. The only one that did this was the self-paced reading study carried out by Carreiras, Garnham, Oakhill and Cain (1996). Differently from the classical case using just one sentence and focused mostly on pronouns (Osterhout, Bersick & McLaughlin, 1997), in their anaphor resolution experiment Carreiras and colleagues manipulated stereotypical role nouns as antecedents in a first sentence, and pronouns as anaphor at the beginning of a second sentence in English [Experiment 1, see example (3)], but in a third sentence in Spanish [Experiment 2, see example (4)]. They measured participants’ self-paced reading times in the sentence segment containing the target. In this case, processing differences on the antecedent part and the pronoun part could be separately evaluated. Participants showed longer self-paced reading times in the second sentence for stereotypical gender mismatch on the pronouns in English and in the first sentence for such mismatch on the role nouns in Spanish respectively.

(3) *The electrician examined the light fitting.*

*He/\*She needed a special attachment to fix it. (p. 644)*

(4) *El<sub>DET/SG/MASC</sub> carpintero/La<sub>DET/SG/FEM</sub> carpintera tomó las medidas para hacer el*

*armario.*

*Era un encargo bastante urgente.*

*El<sub>PRO/SG/MASC</sub>/Ella<sub>PRO/SG/FEM</sub> tenía que terminarlo en el plazo de una semana.*

(p. 648)

*(The male/female carpenter took measurements to make the cupboard.*

*It was a quite urgent order.*

*He/She had to finish it in the space of one week.)*

(examples taken from Carreiras, Garnham, Oakhill & Cain, 1996)

According to the authors, an “assumed” default stereotypical gender was activated immediately after participants encountered a role noun. The time courses for dealing with the activated gender differed due to the distinct morphological gender markings in English and Spanish. For English speakers, the gender activated in the first sentence could only be verified when participants read the pronoun in the second sentence. Contrarily, such verification could be made already when Spanish speakers encountered the noun phrase (determiner + noun) in the first sentence.

To further clarify that the effect was driven by the mismatched stereotypical gender presented in the noun phrases in Spanish, two more experiments (Experiment 3 and 4) were carried out. Particularly in Experiment 4, each first sentence (taken from Experiment 2) was divided into two consecutive displays [i.e., the noun phrase itself and the rest of the sentence, see example (4)] for the self-paced reading procedure. Longer reading times for the first display were found when the gender of a noun phrase mismatched its stereotypical gender (e.g., *la<sub>DET/SG/FEM</sub> carpintera*/the female carpenter, compared to *el<sub>DET/SG/MASC</sub> carpintero*/the male carpenter). Based on these results, the authors concluded that the activation of stereotypical gender is “[...] *an elaborative inference that is not necessary for local cohesion*” (Carreiras, Garnham, Oakhill & Cain, 1996, p. 659) and thus against the minimalist hypothesis (proposed by McKoon & Ratcliff, 1992: inferences are made for the need of local cohesion only. Information that is not addressed in the local structure will not be activated). This conclusion implies that the activation of the default stereotypical gender does not necessarily follow the automatic syntactic processing of gender agreement. On the other hand, the default stereotypical gender could be triggered (or inferred) via lexical access of a role noun regardless of the linguistic form of the text. This position goes against Harris’ (1991) viewpoint that lexical gender of human role nouns can only be accessed through grammatical gender (“[...] *all Spanish nouns have lexical gender, either masculine or feminine (but not neuter). For nouns referring to humans, grammatical gender usually matches biological sex*”, Harris, 1991, p. 36). Thus, stereotypical knowledge indeed also plays an important role for gender processing in Spanish, a grammatical gender language.

As grammaticality was not manipulated in Carreiras, Garnham and Oakhill (1996), the noun phrases with mismatching stereotypical gender were grammatically correct. Thus, how grammatical gender and stereotypical gender interact during the comprehension of Spanish role nouns remains unclear. Therefore, to better understand the processing of stereotypical gender in a grammatical gender language, like Spanish, would help to focus more on (i) the role played by grammatical gender for gender processing on stereotypical role nouns, (ii) how the elaborative inference (e.g., the assumed default stereotypical gender) is activated and (iii) how grammatical and stereotypical gender interact during online processing of stereotypical role nouns.

We took advantage of manipulations involving stereotypical knowledge throughout the present dissertation. We aimed to investigate how stereotypical

knowledge differently interacts with distinct morpho-syntactic gender marking systems. Available studies involving the processing of gender information have been mainly developed in natural gender languages (such as English) and grammatical gender languages (such as German, Italian or Spanish). Very little is known about genderless languages, where, for example, morpho-syntactic gender agreement does not even exist and the gender specificity encoded in the pronouns is different from that in other categories of languages. In the literature, anaphoric pronoun resolution is the most common used language structure for studying the processing of stereotypical gender information. However, such stereotypical processing may be constrained by the linguistic context of anaphor resolution structure, i.e., few words after the antecedent where gender information is processed and consequently retrieved. To directly investigate the online processing of stereotypical gender, the immediate activation of this information should be measured on the role nouns.

Thus, in the present dissertation, the first experiment investigates anaphoric pronoun resolution in a genderless language, i.e., Mandarin Chinese, with a more restricted manipulation on antecedents (definitional role nouns vs. stereotypical role nouns) and pronouns (gender specificity: general 他/*he* vs. specific 她/*she*). By employing the ERP mismatch paradigm, we hope to provide a comparable study parallel for the mapping of the two-stage model and ERP findings in a more comprehensive manner across the three language groups. The second and third experiments focused on immediate activation of stereotypical knowledge on the role nouns in both Spanish (a grammatical gender language) and Mandarin Chinese (a genderless language) to see how distinct morpho-syntactic gender markings in different languages influence the processing of stereotypical knowledge.

### **1.3 The Present Dissertation**

Three experiments (one in Spanish and two in Chinese) were carried out employing the ERP mismatch paradigm. As mentioned in section 1.2, ERP components such as the LAN, the N400 and the P600 can index language processing. In addition to the three components, the P200 component has been reported in studies related to visual attention (Luck & Hillyard, 1994) and is typically associated to the recognition of Chinese characters (graphically, semantically or phonologically) (Liu, Perfetti & Hart, 2003; Lee, et al, 2007; Hsu, Tsai, Lee & Tzeng, 2009; Yum, et al, 2014). Thus, these four ERP components will be the dependent variable throughout this dissertation to see how gender information is processed during language processing. Each experiment is briefly introduced separately in each of the following sections.

#### **1.3.1 Experiment I: Anaphoric Pronoun Resolution in A Language Where Gender is not Morpho-syntactically Encoded**

Regarding the processing of gender information, abundant studies involving anaphoric pronoun resolution are available in grammatical gender languages and natural gender languages but not that many in genderless languages. Although several studies have tried to address this issue in genderless languages (Finnish: Pyykkönen, Hyönä & van Gompel, 2010; Chinese: Wang, Yang & Zhao, 2010; Qiu, Swaab, Chen & Wang, 2012; Xu, Jiang & Zhou, 2013), how definitional gender and stereotypical gender is processed is still unclear, especially when unequal gender specificity is encoded in the pronouns (i.e., 他/*he* generally refers to every individual and 她/*she*

specifically refers to females only). To understand this issue and take advantage of Mandarin Chinese (i.e., a genderless language in which gender is not morpho-syntactically presented and two third person pronouns provide unequal gender specificity), Experiment I is focused on Chinese anaphoric pronoun resolution with manipulations on antecedents (definitional vs. stereotypical) and pronouns (general vs. specific). Sentences with gender congruent and incongruent pronouns are compared. Antecedents with definitional role noun are illustrated in (5) and with stereotypical role noun in (6).

- (5) 在訪談中, 達賴喇嘛表示他自己/\*她自己不會繼續擔任西藏的政教領袖。  
(*In the interview, the **Dalai Lama** expressed **himself**/\***herself** about not continuing to serve as the political religious leader in Tibet.*)
- (6) 比賽前, 那個拳擊手認為他自己/\*她自己很有實力一定可以贏得金牌。  
(*Before the race, **that boxer** considered **himself**/\***herself** very competent for winning the gold medal.*)

### 1.3.2 Experiment II: The Online Processing of Stereotypical Gender in Grammatical Gender Language

Since few studies directly investigated the online processing of stereotypical knowledge on the role nouns, Experiment II is aimed to investigate the activation of stereotypical knowledge in a grammatical gender language (i.e., Spanish) to see how stereotypical gender information interacts with grammatical gender (morpho-syntactic gender).

Carreiras and colleagues' study (1996) focused on stereotypicality in grammatically correct sentences (e.g.,  $lo_{DET/MASC/PL}$   $minerOS_{MASC/PL}$ /the male miners vs.  $la_{DET/FEM/PL}$   $minera_{FEM/PL}$ /the female miners) and grammaticality was not manipulated on the role nouns (e.g.,  $lo_{DET/MASC/PL}$   $minerOS_{MASC/PL}$ /the male miners vs.  $*la_{DET/FEM/PL}$   $minerOS_{MASC/PL}$ /\*the male miners). How stereotypical gender is influenced by grammatical gender was not investigated. Therefore, the interaction between stereotypical gender and grammatical gender within the same noun phrase (i.e., determiner + role noun) during sentence processing is the focus in Experiment II.

### 1.3.3 Experiment III: The Online Processing of Stereotypical Gender in Genderless Language

As mentioned in section 1.1.2, Mandarin Chinese is categorized as a genderless language for the lack of morpho-syntactic gender morphemes both on the role nouns and pronouns. No morpho-syntactic gender agreement is required between any related words either. The way gender is expressed in Chinese is by means of semantic gender modifiers/adjectives (Farris, 1988; Moser, 1997) and sometimes by the semantic radical encoded in a character (e.g., the semantic radical in Chinese third person pronouns: radical '人' (/rén/human) for 他/*he* and radical '女' (/nǚ/woman) for 她/*she*, though both pronounced as /tā/). All Chinese role nouns are generic, but can still be biased towards to one gender (i.e., the cover category or stereotypical gender, see section 4.1 for details) and this generic role noun represents the default stereotypical gender of that role noun (e.g., 搬運工人/*porter*, male-biased). To clearly indicate one's gender, a semantic gender modifiers/adjective, 男(的) (*male*) or 女(的) (*female*) could be added before a role noun to form a gender specified role noun [e.g., “男(的)”

搬運工人/“male” porter or “女(的)” 搬運工人/“female” porter. The 的/de is a nominalization particle depending on the literary style in the context (Li & Thompson, 1981).

Different from Indo-European languages, Chinese gender marking system is semantic (Farris, 1988; Moser, 1997). Following Experiment II, Experiment III is aimed to look at the activation of stereotypical knowledge in a genderless language to see how such activation on the role nouns (and the interaction between stereotypical and “formal” gender information) differs by distinct gender marking systems (e.g., with/without morpho-syntactic grammatical gender among others). Sentences like (7) are compared.

- (7) 大樓門口，一些搬運工人/男的搬運工人/\*女的搬運工人正在將貨物搬上車。  
*(In front of the building, some porters/male porters/\*female porters are loading the goods on the truck.)*

### 1.3.4 Summary of Chapter 1

Gender is a basic feature related to every animate entity and this feature is incorporated into language use. However, how gender information is expressed in the linguistic code differs across languages. Based on the grammatical features of the coding, languages are categorized into three groups: grammatical gender languages, natural gender languages and genderless languages.

This dissertation is focused on how gender information is processed in languages with distinct morpho-syntactic gender marking systems, looking at anaphoric pronoun resolution in Mandarin Chinese (a genderless language) and the online activation of stereotypical knowledge on the role nouns both in Spanish (a grammatical gender language) and Mandarin Chinese.

The detailed literature review, experimental methods, results and the interpretation of the findings are presented in each chapter. Experiment I is presented in Chapter 2, Experiment II is presented in Chapter 3, and Experiment III is presented in Chapter 4. The overall discussion is developed in Chapter 5.

## **CHAPTER 2**

### **Experiment I: Pronoun Resolution in A Language Where Gender Is not Morpho-syntactically Encoded**

## 2.1 Introduction<sup>1</sup>

Anaphor resolution involves linking a given anaphor to a previously-mentioned antecedent in a sentence context, while interpreting both as related to the same discourse-level entity. Most studies related to the effects of gender information on anaphor resolution have proposed that (morpho)-syntactic rules constrain anaphor resolution (Carreiras, Garnham & Oakhill, 1993; Garnham, Oakhill, Ehrlich & Carreiras, 1995; Carreiras, Garnham, Oakhill & Cain, 1996; Osterhout, Bersick & McLaughlin, 1997; Kennison & Trofe, 2003; Sturt, 2003; Duffy & Keir, 2004; Kreiner, Mohr, Kessler & Garrod, 2009; Esaulova, Reali & von Stockhausen, 2014). From a theoretical perspective, anaphoric processing is considered to involve two stages of processing (Garrod & Sanford, 1994; Garrod & Terras, 2000). A bonding stage occurs first, for the purpose of searching for a best-fit referent among possible candidates (i.e. the antecedents) based on information related to gender, number and syntactic rules. A resolution stage subsequently accomplishes the interpretation of the anaphor, taking into account world knowledge and contextual information.

There is an open debate, however, about the degree to which lexical-semantic and syntactic cues are employed to resolve the anaphor. Carreiras, Garnham, Oakhill and Cain (1996) studied what linguistic information constrains pronoun resolution processing, by employing stereotypical role nouns as antecedents. They measured English and Spanish speakers' self-paced reading times and found cross-linguistic differences for the different morphological gender marking systems in English and Spanish. According to these authors, as soon as the mismatch of stereotypical gender is detected by the participants (i.e., on the pronouns encountered after stereotypical role nouns in English or on stereotypical role nouns at the beginning of a sentence in Spanish), it immediately influences the on-going processing, resulting in different patterns during pronoun resolution. Eye-tracking studies using stereotypical role nouns as antecedents (Sturt, 2003, in English; Esaulova, Reali & von Stockhausen, 2014, in German) have supported a two-stage model of anaphor resolution and these authors interpreted their findings as evidence for the syntactic constraints employed in resolving the link between an anaphor and its antecedent at the initial stage. Esaulova and colleagues (2014) concluded: "*anaphor resolution [...] seems to depend above all on the rules of grammatical agreement in the context of overlapping gender cues*" (p. 798).

Osterhout, Bersick and McLaughlin (1997) carried out an ERP study in English, in which the gender type of the antecedent (definitional vs. stereotypical) and the antecedent-reflexive pronoun gender congruity were manipulated (see also Kreiner, Mohr, Kessler & Garrod, 2009 for an ERP study and Kreiner, Sturt & Garrod, 2008 for an eye-tracking study). While definitional role nouns had a definite gender (e.g., mother/father), gender in stereotypical role nouns was inferred based on world knowledge (i.e., a role noun that could refer to two genders but is biased towards one, e.g., electrician is male-biased and beautician is female-biased). The results showed similar P600 effects for antecedent-reflexive pronoun gender mismatches in conditions of both definitional (*mother - himself*) and stereotypical (*nurse - himself*) role nouns. These authors concluded that gender information was grammatically encoded even for stereotypical role nouns. Based on their interpretation, it would follow that anaphor resolution correlates with a single ERP component, the P600, and

<sup>1</sup> The present experiment is published in *Frontiers in Psychology* and the reference is as follows: Su, J.-J., Molinaro, N., Tsai, P.-S., Wu, D.H., & Carreiras, M. (2016). When "he" can also be "she": An ERP study of reflexive pronoun resolution in Mandarin Chinese, *Frontiers in Psychology*, 7, 151.



engages a process that is syntactic in nature rather than semantic/pragmatic. However, Nieuwland and Van Berkum (2006) found an N400 effect for pronouns with antecedents of the same gender as compared to those with antecedents of different gender in a Dutch study and attributed the results to differences in the contextual bias that would modulate the N400 effect (used to index semantic/context related processing) during anaphor resolution.

Further results obtained from German studies are heterogeneous. Schmitt, Lamers and Münte (2002) investigated ERP responses to pronouns related to biological (definitional) and grammatical gender entities. The authors claimed that anaphor resolution is basically syntactically driven (P600 effect reported) but can interact with semantic information in the N400 time interval. Irmen, Holt and Weisbrod (2010) focused on the link between antecedent stereotypical gender and anaphor lexical-semantic gender (these men/women/people) in German. They reported an N400 stereotypical gender effect and a P600 effect for anaphor mismatch with the antecedent's stereotypical gender. These authors interpreted their findings as supporting the two-stage model: while stereotypical gender information is collected in the bonding stage, the resolution stage represents integration, driven by either lexical-semantic mismatch or syntactic violation on the anaphors.

Results from Chinese ERP studies also reveal somewhat different patterns. Qiu and colleagues (2012) manipulated the distance and gender congruity between antecedent and pronoun in Chinese sentences. N400 and P600 mismatch effects were found respectively for short and long distance manipulations. These authors claimed that the processing of gender information in Chinese pronoun resolution is more 'semantics-based' when the pronoun is closer to the antecedent and the semantic nature of this representation decays as the distance increases. Xu, Jiang and Zhou (2013) (Experiment 1, singular antecedent) also manipulated gender congruity between antecedent and pronoun across clause boundaries (i.e., long-distance dependency). Only P600 effects were reported for gender mismatches. However, the authors did not interpret this P600 effect as reflecting purely syntactic processing, but either semantics-based processing -computation of the semantic relationship between antecedent and anaphor- or general integration difficulty resulting from conflicts on gender during Chinese anaphor resolution. It is important to note, however, that in these two Chinese studies, the gender specificity of the pronouns (i.e., 他/*he* & 她/*she*) was considered to be symmetrical, as is the case in morphological gender languages. However, co-reference processing between the anaphor and the antecedent could in fact differ in Chinese, due to the asymmetry of gender specificity encoded in the Chinese characters for these pronouns.

Indeed, the processing of gender information has been investigated by employing the co-indexation structure of anaphor resolution mainly in languages in which the morpho-syntactic gender marking on anaphor<sup>2</sup> is "symmetrically" expressed, i.e., "he/himself" is used specifically for male antecedents and "she/herself" exclusively refers to female antecedents. However, in a language without inflectional morphological gender markings, such as Chinese<sup>3</sup>, the gender specificity is not

<sup>2</sup> The term anaphor here refers to all kinds of anaphors in general that could refer back to an antecedent in previous text.

<sup>3</sup> According to Packard (2000) on Chinese word formation, the basic orthographic unit in Chinese written words is the character and generally one character represents one morpheme. From this point of view, Chinese has morphological gender markings, which are represented by individual characters/morphemes instead of inflections on a root. The term "morphological markings/cues", however, usually refers to inflectional morphology.

symmetrically encoded in the written forms of the pronouns. In the spoken language, the third person singular pronoun is pronounced the same, /tā/, for both genders and the gender of a pronoun is inferred based on the context. Gender distinction is thus made only in written Chinese. Although the characters for the male and female pronouns share the same phonological component “也”, they differ in their semantic radicals<sup>4</sup>. According to the web-based Dictionary of Chinese Character Variants established by the Ministry of Education in Taiwan, the character 他/*he*, which contains the semantic radical ‘人’ (/rén/*human*) is the “third person pronoun, refers to a third person” and the character 她/*she*, pronounced as /tā/, containing the semantic radical ‘女’ (/nǚ/*woman/female*) is the “female third person pronoun<sup>5</sup>.” When reading Chinese, therefore, during anaphor resolution it could be the case that the relevance of the gender information provided by an antecedent may differ depending on the extent to which gender information is specifically presented or not in the anaphoric pronoun. The processing of the anaphor may also differ when the antecedent’s gender is either definitional, with a clear gender, or stereotypical, where gender can only be inferred. This critical feature (i.e., asymmetry of gender specificity in Chinese pronouns) has rarely been tested in previous Chinese studies and could result in distinct patterns of processing of gender information.

We illustrate below why the gender specificity encoded in the pronouns is not likely to be symmetrical, from three different perspectives: the historical background of the characters, the difference in their semantic radicals, and the usage of 他/*he* as a default. First, from a historical point of view, 她/*she* was only recently proposed as the third person singular female pronoun by a linguist, Ban-Nong Liu, in 1921 for the convenience of translation from western languages (Ling, 1989; Chang, 1994; Hua, 2012)<sup>6</sup>. Previous to this, the default 他/*he* was used in written Chinese. The use of the character 她/*she* was not generally accepted at first and even now its appearance and necessity remains controversial (Chang, 1994; Moser, 1997; Wang, 2010; Hua, 2012). In recent years, due to campaigns for gender equality, the two pronouns have been gradually differentiated but the use of 她/*she* is still not compulsory for female antecedents (Peng, 2009).

Secondly, the semantic radicals encoded in the two pronouns play a critical role in bringing out the gender specificity. In Chinese, the radical 人/rén/ means *human* and the radical 女/nǚ/ means *woman/female* and these different semantic radicals make the characters orthographically distinct from each other. Increasing evidence has shown that Chinese speakers rely very much on sub-lexical units -semantic radicals and phonological components- during text comprehension (Perfetti & Zhang, 1991; Feldman & Siok, 1999; Ho, Ng & Ng, 2003; Liu, Perfetti & Hart, 2003; Ding, Peng & Taft, 2004; Lee, et al., 2007; Tsang & Chen, 2009; Hsu, Tsai, Lee & Tzeng, 2009). A study carried out by Cherng, Chang and Chen (2009), which explored whether Chinese script reflects negative attitudes towards women (whether characters

<sup>4</sup> Typically, a written Chinese character represents one morpheme and one syllable and is composed of two parts: a semantic radical indicating the meaning and a phonological component that provides information about the pronunciation.

<sup>5</sup> The plural is formed by adding another character, 們/men/, which is a collective marker, to either of these pronouns to form 他們/tā men/*they* (male) or 她們/tā men/*they* (female) (Li & Shi, 2000).

<sup>6</sup> The character 她/*she* was recorded as first appearing in 玉篇 (Article of Jade, a manuscript written in the Yuan Dynasty), but the pronunciation and meaning is identical to 姐/jiě/ meaning elder sister or ladies (Lin, 1989; Hua, 2012; Dictionary of Chinese character variants established by the Ministry of Education in Taiwan). The Yuan Dynasty was from 1271 A.D.-1368 A.D.

containing the semantic radical for “woman” have a more negative valence), found no evidence of this in Chinese speakers’ perception of gender-based characters. They reported no negative attitudes towards characters containing the semantic radical for “woman” (the meaning conveyed by the characters was rated by participants as positive, negative or neutral). However, while characters containing the radical 子 /zǐ/son and radical 女/nǚ/woman were rated as positive, characters with the radical 人 /rén/human were rated as neutral. These results may imply differences in the mental representations of the gendered semantic radicals, especially when they appear in pronouns denoting different gender specificity.

Third, from an empirical point of view, the asymmetry on gender specificity presented by the pronouns is reflected in the usage of 他/*he* as a default. Wu and Liang (2008) analyzed 150 news items taken from the Academia Sinica Balanced Corpus (ASBC), for a rule-based corpus analysis of Chinese pronominal anaphor resolution. The results showed an error rate of .21 for gender mismatches between 他/*he* and female antecedents. The authors attributed this relatively high rate of mismatch to the use of 他/*he* as default in Chinese written text. Different learning sequences of the two pronouns at school also contribute to the tendency to use 他/*he* as a default. A textbook analysis carried out by Huang and Luh (2012) in Taiwan reported that while children learn 他/*he* in the first year of elementary school, 她/*she* is learnt in the second year. In some textbook articles, 他/*he* is used to refer to female antecedents before and even after the pronoun 她/*she* is learnt. Word frequencies of the pronouns reported by Academia Sinica (Word List with Accumulated Word Frequency in Sinica Corpus<sup>7</sup>) correspond to such usage trends in reality (see Table 1).

Data from these three perspectives thus clearly indicate that gender specificity in the characters for Chinese third person pronouns is non-symmetrical. Investigating this asymmetry can shed light not only on the processing of gender information in written Chinese but also on the general processing of pronouns in text.

**Table 1. Word Frequencies for the Two Chinese Pronouns Extracted from Word List with Cumulated Word Frequency in Sinica Corpus, Academia Sinica, Taiwan.**

	他/ <i>he</i>	她/ <i>she</i>
Cumulative Word Frequency for Modern Chinese words (based on the corpus size of 5 million words)	29938	10755
Cumulative Word Frequency for Pre-modern Chinese Corpus	37259	2
Cumulative Word Frequency for Old Chinese Corpus	36	Word not found

In the present study, we investigate how the asymmetry of gender specificity interacts with antecedent noun type in which the biological gender is differently inferred during anaphor resolution in written Chinese. The experiment had a 2 x 2 x 2 design, with three factors manipulated: antecedent gender type (definitional vs. stereotypical), reflexive pronoun gender specificity (default vs. specific), and gender congruity between the reflexive pronoun and antecedent (congruent vs. incongruent).

<sup>7</sup> Word List with Accumulated Word Frequency in Sinica Corpus (2005). Available online at: [http://elearning.ling.sinica.edu.tw/eng\\_teaching\\_index.html](http://elearning.ling.sinica.edu.tw/eng_teaching_index.html) (Accessed April 30, 2015).

First, antecedent gender type was manipulated. For definitional gender, because most female definitional role nouns in Chinese carry a 女/nǚ/woman radical in the same position as 她/she, (such as 媽/mā/mother, 嬸/shěn/aunt, & 姐/jiě/elder sister, Tang, 1988), proper names were used, to limit priming effects due to the presence of the same semantic radical in antecedent and reflexive pronouns (see Feldman & Siok, 1999; Ding, Peng & Taft, 2004). So, proper names such as 達賴喇嘛/Dalai Lama or 呂秀蓮/Annette Lu were used as antecedents for definitional gender. Stereotypical role nouns (e.g., 拳擊手/quán jī shǒu/boxer, male-biased; or 啦啦隊長/lā lā dui zhǎng/cheerleader, female-biased) were used as antecedents for the stereotypical gender condition.

Second, gender specificity (default vs. specific) was manipulated on reflexive pronouns instead of personal pronouns. In Chinese, the use of pronouns is not constrained by the local structure (Principle B, Chomsky, 1981: pronouns cannot co-refer with antecedents in a local clause). A reflexive term, 自己/zì jǐ/self is allowed to appear after a pronoun to make it a reflexive pronoun co-referential to the subject in the same clause (Principle A, Chomsky, 1981), (see Li & Thompson, 1981). For instance, in (8a) 她/she could refer to the teacher or another female. In (8b), the reflexive, 自己/zì jǐ/self, helps to make the pronoun unambiguously co-referential to the previously mentioned animate antecedent (i.e., Mary) in the same clause (i.e., local binding of 自己/zì jǐ/self, see Jäger, Engelmann & Vasishth, 2015). Therefore, to avoid any confusion in co-reference between an anaphor and its antecedent, the reflexive 自己/zì jǐ/self was added after the third person pronouns to form the third person reflexive pronouns (他自己/tā zì jǐ/himself, default; and 她自己/tā zì jǐ/herself, specific).

- (8) a. 這個老師覺得 Mary 不喜歡 她。  
 /zhè gè/ /lǎo shī/ /jiào dé/ /Mary/ /bù xǐ huān/ /tā/  
 (This teacher thinks that Mary doesn't like **her**.)
- b. 這個老師覺得 Mary 不喜歡 她自己。  
 /zhè gè/ /lǎo shī/ /jiào dé/ /Mary/ /bù xǐ huān/ /tā zì jǐ/  
 (This teacher thinks that Mary doesn't like **herself**.)

Third, we manipulated gender congruity between a reflexive pronoun and its antecedent (congruent vs. incongruent). It should be noted that when 他自己/himself appears after a female/female-biased antecedent, the sentence might still be acceptable in Chinese because 他自己/himself can be used as a general term referring to both genders, even though we consider this as a mismatch in the data analysis. An ERP mismatch paradigm was employed.

In ERP studies related to anaphor resolution, three major correlates have been discussed (see Callahan, 2008): the LAN (Left Anterior Negativity), the N400 and the P600. The LAN is a negative-going wave mostly observed at left anterior scalp electrodes from around 250 ms to 500 ms post target word onset. The LAN is related to automatic parsing involving morphosyntax and rule-based decomposition processes (Hahne & Friederici, 2002; Barber & Carreiras, 2005; Hagoort, 2009; Molinaro et al., 2008a, 2008b, 2011). The N400 is a negative-going wave peaking around 400 ms after the onset of the target word, mainly recorded from electrodes in the centroparietal scalp regions (Kutas & Hillyard, 1980; Molinaro et al. 2010). The N400 is thought to represent context-dependent, lexico-semantic processing of a given

stimulus. Its amplitude can be modulated depending on the lexical properties of single words and, at the sentence level, the anticipation/contextual semantic fit of a word with the previous context (Kutas & Federmeier, 2011, for a review). The P600 effect is a positive-going wave observed approximately 500-700 ms after target word onset, with centroparietal scalp distribution. This component was initially reported as correlating with (morpho-) syntactic violations (Osterhout & Holcomb, 1992). Recently, however, the interpretation of the P600 effect has been extended to reflect more general (also semantic) integration difficulties during sentence processing (Münte et al., 1998; Kaan, Harris, Gibson & Holcomb, 2000; Kaan & Swaab, 2003; Carreiras, Salillas & Barber, 2004; Van Herten, Kolk & Chwilla, 2005; Van de Meerendonk, Kolk, Chwilla & Vissers, 2009; Brouwer, Fitz & Hoeks, 2012; Molinaro, Carreiras & Duñabeitia, 2012). In addition to these three components, the P200 component reflects a wide range of attention-related feature analysis including color, orientation and size of a feature (Luck & Hillyard, 1994). It is reported in studies related to word frequency (Dambacher, Kliegl, Hofmann & Jacobs, 2006), syllable frequency/structure (Barber, Vergara & Carreiras, 2004; Carreiras, Vergara & Barber, 2005) and Chinese character recognition (graphic, semantic or phonological) (Liu, Perfetti & Hart, 2003; Lee, et al, 2007; Hsu, Tsai, Lee & Tzeng, 2009; Yum, et al, 2014). Critically, the P200 component has recently been reported in Chinese discourse inference (Hung & Schumacher, 2012; 2014) and is considered to be related to a certain expectation driven by the context. It is reported as distributed mostly in the anterior region; the more positive the P200 amplitude, the less familiar, lower frequency and less expected the target word.

Based on previous ERP studies and the two-stage model of anaphor resolution (Garrod & Sanford, 1994), interactions among antecedent gender explicitness, reflexive pronoun gender specificity and gender congruity are thus expected mainly in the N400 or P600 time intervals. Since the pronoun 他/*he* could be considered as a default pronoun (that can refer to both male and female antecedents) (Wu & Liang, 2008) and supported by the word frequencies of both pronouns in the Sinica Corpus, two hypotheses are possible about how the default pronoun is perceived for gender. First, if the 他/*he* is recognized as male-biased, a gender mismatch N400 or P600 effect is predicted for 他自己/*himself* that appeared after a female antecedent. Second, if the 他/*he* is understood as equally applicable to both genders, no gender mismatch effects are expected for 他自己/*himself* following female or female-biased antecedents (e.g., 呂秀蓮/*Annette Lu* or 啦啦隊長/*cheerleader*). Whether or not the second hypothesis results in a mismatch effect is considered a key result in the present study, to evaluate participants' sensitivity to the gender asymmetry of the reflexive pronouns. On the other hand, stronger gender mismatch effects are expected for the female specific reflexive pronoun, 她自己/*herself*, because this contrast (male/male-biased antecedent vs. specific female reflexive) determine an unambiguous gender incongruity. This contrast is also critical in evaluating how the specific reflexive pronoun is processed depending on the gender type of the antecedent. Both N400 and P600 effects are expected in this "pure" mismatch gender contrast, reflecting semantic/pragmatic processing costs (Osterhout, Bersick & McLaughlin, 1997; Callahan, 2008; Tsai et al., 2009; Irmén, Holt & Weisbrod, 2010; Molinaro et al., 2012; Qiu, Swaab, Chen & Wang, 2012; Xu, Jiang & Zhou, 2013). For the relatively less familiar and lower frequency orthographic form (i.e., the specific reflexive pronoun 她自己/*herself*), the attention-related P200 effect (Luck & Hillyard, 1994; Liu, Perfetti & Hart, 2003; Lee, Liu & Tsai, 2012; Hung & Schumacher, 2012; 2014)

and the lexical frequency-related N400 effect (Kutas & Federmeier, 2011) are also expected. Specifically, we are interested in the time course of the ERP effects, to see when and how the linguistic sources of gender information denoted by the reflexive pronouns interact with antecedent gender type and gender congruity during reflexive pronoun resolution.

## **2.2 Materials and Methods**

### **2.2.1 Participants**

Forty native Chinese speakers (20 males, mean age: 21.8, aged 20-36 years) were recruited from the National Central University, Taiwan and were paid for their participation. They were healthy, right-handed, with normal or corrected-to-normal vision and reported no neurological or psychiatric history. The design and execution of the experiment conformed to the ethical regulations of the Institute of Cognitive Neuroscience at National Central University in Taiwan, which are equivalent to international standards. Informed consent was obtained from each participant.

### **2.2.2 Materials**

Proper names of eighty celebrities (40 males, see Appendix A) were selected from the news, based on Google search from July to September 2011. The names chosen had a number of occurrences larger than 300,000. Eighty sentences related to the 40 male and 40 female celebrities were used. Half of the 40 sentences containing the male/female proper names had a congruent third person reflexive pronoun and the other half had an incongruent one. These sentences were the experimental materials for the condition of antecedents with definitional gender.

The stereotypical role nouns were selected based on the results of a questionnaire containing three hundred forty-eight generic role nouns (e.g. 拳擊手/*quán jī shǒu/boxer*, 學生/*xué shēng/student*, or 啦啦隊長/*lā lā duì zhǎng/cheerleader*). Participants were asked to mark on an 11-point Likert scale from 0% to 100% with the percentage of ten as the scale interval. The point of 0%-male/100%-female was always on the left and 100%-male/0%-female was always on the right. Fifty-seven college students (14 males, mean age: 21.58) from Tsing Hua University and Sun Yat-Sen University filled in this questionnaire. The forty most male-biased role nouns and 40 most female-biased ones were used as antecedents with stereotypical gender in 80 sentences (see Appendix B). Half of the male-biased/female-biased antecedents were associated with a congruent third person reflexive pronoun and the other half with an incongruent one in sentences. The sentence structure for all materials was constructed as short-distance [i.e., the antecedent and the reflexive pronoun were in the same clause. See (9) and (10)]. At the beginning of each sentence, a segment denoting a time, location or circumstance appeared. This was followed by a clause with S+V+O structure. The subject (i.e., antecedent) was presented by means of a proper name or stereotypical role noun. The object (i.e., anaphor) was the reflexive pronoun referring back to the subject. The target word was always the reflexive pronoun in each sentence, located in the fourth, fifth or sixth position and never appeared at the end of the sentence.

- (9) 記者會上，呂秀蓮表示 **她自己** 會繼續支持反核。  
/jì zhě huì shàng/, /Annette Lu/ /biǎo shì/ /tā zì jǐ/ /huì/ /jì xù/ /zhī chí/ /fǎn hé/  
(In the press conference, Annette Lu expressed **herself** about continuing to support the anti-nuclear movement.)
- (10) 在訪談中，達賴喇嘛表示 **他自己** 不會繼續擔任西藏的政教領袖。  
/zài fǎng tán zhōng/, /Dalai Lama/ /biǎo shì/ /tā zì jǐ/ /bú huì/ /jì xù/ /dān rèn/  
/xī cáng de/ /zhèng jiāo lǐng xiù/  
(In the interview, the Dalai Lama expressed **himself** about not continuing to serve as the political religious leader in Tibet.)

An additional 80 filler sentences were created. The manipulations on critical words were focused on whether their semantic meaning could fit into the sentence or not (40 semantic match vs. 40 semantic mismatch). Data from these 80 sentences were not included in the data analysis. In total, 240 sentences were employed in this study. Thirty percent of the sentences were accompanied by comprehension yes/no questions (i.e., 72 questions) to evaluate participants' understanding of the sentences. The questions were related to the description of the main character and never related to any gender information of our interest. List 1 contained all the 240 sentences mentioned above. For counter-balancing purposes, a second list was created. List 2 contained the same 240 sentences as in List 1 but with all the target words presenting the opposite manipulation (see Table 2 for examples of materials).

**Table 2. Example Sentences Used in the Experiment**

	Default reflexive pronoun, 他自己/ <i>himself</i>	Specific reflexive pronoun, 她自己/ <i>herself</i>
<b>Antecedent with definitional gender</b>		
<b>Gender congruent</b>	在訪談中, 達賴喇嘛表示 他自己不會繼續擔任西藏的政教領袖。(20 sentences) ( <i>In the interview, the <b>Dalai Lama</b> expressed <b>himself</b> about not continuing to serve as the political religious leader in Tibet.</i> )	記者會上, 呂秀蓮表示 她自己會繼續支持反核。(20 sentences) ( <i>In the press conference, <b>Annette Lu</b> expressed <b>herself</b> about continuing to support the anti-nuclear movement.</i> )
<b>Gender incongruent</b>	*記者會上, 呂秀蓮表示 他自己會繼續支持反核。(20 sentences) (* <i>In the press conference, <b>Annette Lu</b> expressed <b>himself</b> about continuing to support the anti-nuclear movement.</i> )	*在訪談中, 達賴喇嘛表示 她自己不會繼續擔任西藏的政教領袖。(20 sentences) (* <i>In the interview, the <b>Dalai Lama</b> expressed <b>herself</b> about not continuing to serve as the political religious leader in Tibet.</i> )
<b>Antecedent with stereotypical gender</b>		
<b>Gender congruent</b>	比賽前, 那個拳擊手認為 他自己很有實力一定可以贏得金牌。(20 sentences) ( <i>Before the race, <b>that boxer</b> considered <b>himself</b> very competent for winning the gold medal.</i> )	表演後, 那個啦啦隊長評論 她自己覺得表現不錯。(20 sentences) ( <i>After the performance, <b>the cheerleader</b> commented on <b>herself</b> for doing not badly.</i> )
<b>Gender incongruent</b>	*表演後, 那個啦啦隊長評論 他自己覺得表現不錯。(20 sentences) (* <i>After the performance, <b>the cheerleader</b> commented on <b>himself</b> for doing not badly.</i> )	*比賽前, 那個拳擊手認為 她自己很有實力一定可以贏得金牌。(20 sentences) (* <i>Before the race, <b>that boxer</b> considered <b>herself</b> very competent for winning the gold medal.</i> )
<b>Fillers</b>	<b>Semantic match</b> 每天早上我搭公車去學校上課。(40 sentences) ( <i>Every morning, I take <b>the bus</b> to school.</i> )	<b>Semantic mismatch</b> *每天早上我搭手術去學校上課。(40 sentences) (* <i>Every morning, I take <b>the surgery</b> to school</i> )

### 2.2.3 Procedure

All the stimuli were presented in white letters on a black background. Each trial began with a fixation point “+” at the center of the screen for 500 ms, followed by a blank screen for 400 ms. Each word was presented for 400 ms followed by a blank screen for 400 ms. According to some previous studies (Jiang & Zhou, 2009; Ye et al., 2007), 400 ms (word) + 400 ms (blank) word presentation is natural and comfortable for Chinese readers. A variable inter-trial time interval (from 1700 ms to 3000 ms) appeared after each sentence.

Participants were comfortably seated in a sound-attenuated cubicle and were instructed to read each sentence silently and carefully. Their task was to answer yes/no comprehension questions by pressing one of the pre-designated buttons (“J” for ‘Yes’ and “F” for ‘No’). A practice session with 12 trials was conducted before the main experiment. The main experiment was arranged in six blocks with five breaks. Each block contained 40 sentences. The 240 sentences were randomly presented, differently for each participant.



## 2.2.4 Data Recoding and Analysis

Continuous EEG data (SynAmps2, NeuroScan) were acquired from 32 active electrodes mounted in a 66-channel Quick Cap. Electrodes were positioned according to the 10-20 system. The impedance was kept below  $5\Omega$  in each electrode. The sampling rate (A/D) was 500 Hz. The on-line reference electrode was set to be the left mastoid (M1) and we also recorded the right mastoid (M2). The signals were amplified with a bandpass of 0.05-100 Hz. The ground electrode was set between FPZ and FZ. HEOGs were placed at the outer canthi of the eyes and VEOGs were placed above and below the left eye in a bipolar montage.

The EEG raw data were re-referenced to the average activity of M1 and M2. The signal was bandpass filtered between 0.1 Hz and 30 Hz. Epochs of interest were from  $-100$  ms before the onset of the target word to 1000 ms after stimulus onset. Baseline correction was set from  $-100$  ms to the onset of the target words. Trials with artifacts, such as eye blinks or saccades, or with activity exceeding  $\pm 120 \mu\text{V}$  were rejected. As a result, 5.6% of the trials were removed due to artifact rejection.

Except for trials with artifacts, all the correctly and incorrectly judged trials were included for statistical analysis. Data analysis focused on the mean voltage of each electrode within a time interval of interest after the onset of the target words in each participant. Based on the findings from previous ERP studies, four components, P200, LAN, N400 and P600 are used to index the processing correlates. The time intervals were chosen based on visual inspection of the averaged wave patterns.

Repeated measures ANOVAs were separately employed for electrodes in the midline region (anterior: average activity of Fz and FCz; central: Cz and CPz; and posterior: Pz and Oz) and in the lateral scalp (left anterior: average activity of Fp1, F3, F7 and FT7; left central: FC3, C3, CP3 and T7; left posterior: TP7, P3, P7 and O1; right anterior: Fp2, F4, F8 and FT8; right central: FC4, C5, CP4 and T8; and right posterior: TP8, P4, P8 and O2). A four-way repeated measure ANOVA was employed for the midline region considering *antecedent gender type* (definitional vs. stereotypical), *reflexive pronoun gender specificity* (default vs. specific), *gender congruity* (congruent vs. incongruent), plus the *latitude* topographical factor (anterior, central and posterior). For electrodes in the lateral scalp, a five-way repeated measures ANOVA was used: *latitude* and *lateral scalp* (left vs. right) were the topographical factors added to the three main factors. The Greenhouse-Geisser corrected *p*-value was used if the degree of freedom was larger than one. For interactions among the experimental factors and/or topographic factors, planned paired *t*-tests (with FDR adjusted *p*-value) were carried out mainly focused on the comparison of gender congruity respectively in each two levels of the main factors and/or separately in the topographic region to look for the location of the effect. Non-significant effects obtained from the planned paired *t*-tests following significant interactions are not reported in the data analysis.

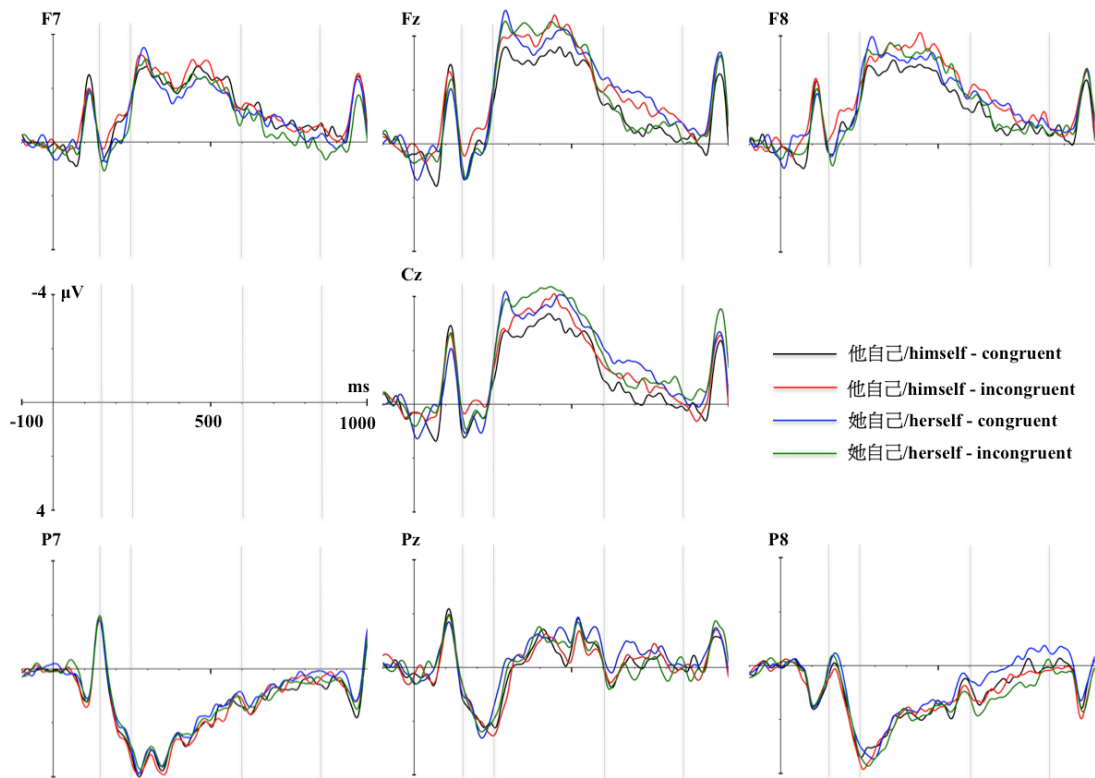
## 2.3 Results

### Comprehension questions

The average of participants' accuracy in the comprehension questions was 93%, ranging from 85% to 99%, showing that participants understood very well the sentences they read.

## ERPs on the reflexive pronouns

Figure 1 reports the grand average of the ERPs elicited by the two reflexive pronouns, taking into account gender congruity. Based on visual inspection of the overall ERP results (and supporting evidence in the literature), the time interval for the analysis of the P200 component was set as 150–250 ms after the onset of stimuli, that for the N400 component was 250–600, and that for the P600 component was 600–800. The repeated measures ANOVAs on the earlier time intervals did not show any statistically reliable effects (i.e., baseline correction: –100–0 ms; 0–150 ms).



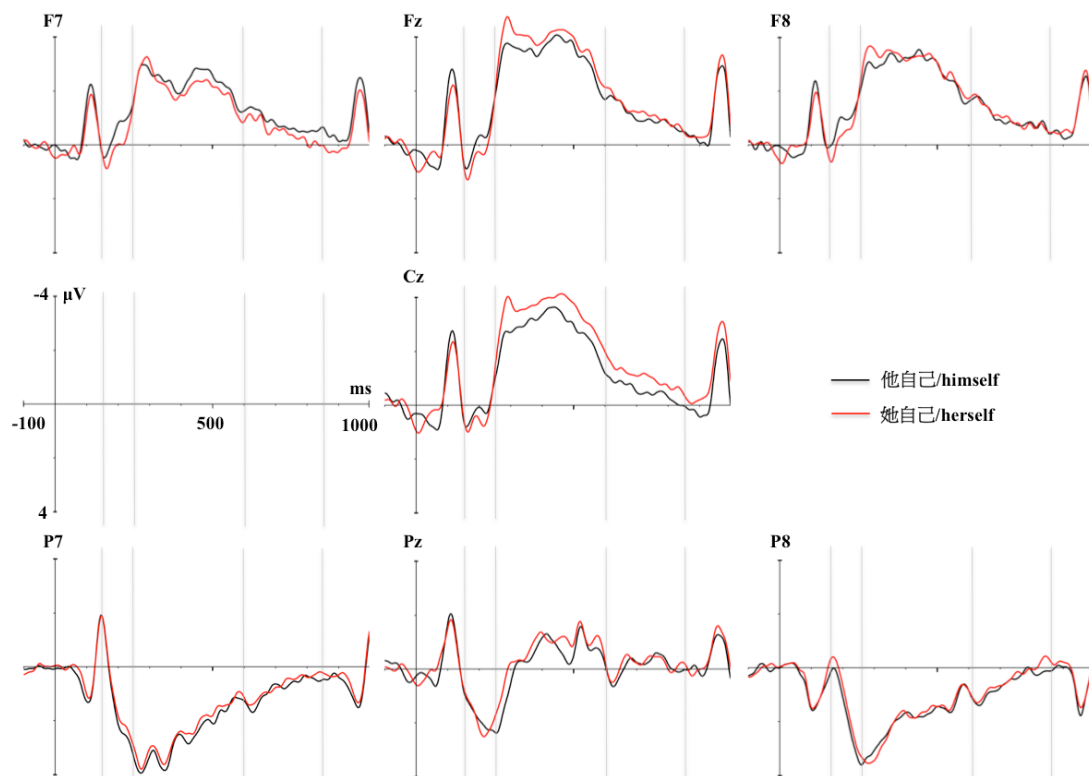
**Figure 1. Overall averaged brain activities of reflexive pronoun gender specificity (general vs. specific) by gender congruity (congruent vs. incongruent) in the representative electrodes.**

### 2.3.1 ERPs on the Reflexive Pronouns at the P200 Time Interval: 150 ms to 250 ms

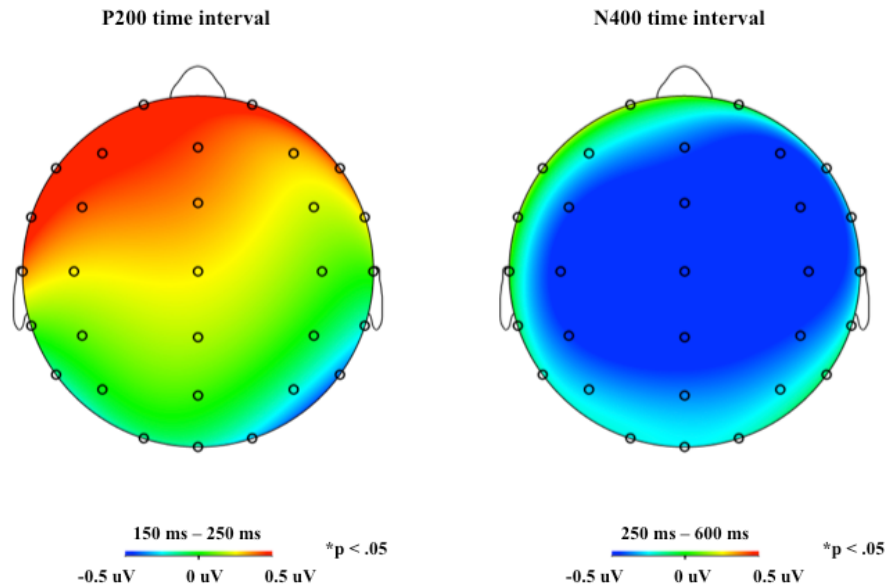
The repeated measures ANOVA on the amplitude of the evoked activity for electrodes in the midline region showed an interaction among reflexive pronoun specificity, gender congruity and latitude [ $F(2, 78) = 4.43, p = .031$ ]. The effect was located in the mid-anterior region for 他自己/himself [congruent:  $0.35 \mu\text{V}$ ; incongruent:  $-0.33 \mu\text{V}$ ;  $t(39) = 2.02, p = .050$ ] and no such effect was found for 她自己/herself (see Figure 1).

The repeated measures ANOVA for electrodes in the lateral scalp regions showed an interaction between *reflexive pronoun specificity* and *latitude* [ $F(2, 78) = 9.34, p = .002$ ] and an interaction among *reflexive pronoun specificity*, *gender congruity* and *latitude* [ $F(2, 78) = 6.18, p = .010$ ]. For the former interaction, the paired *t*-tests showed significant differences in the anterior region in which 她自己/*herself* was more positive than 他自己/*himself* [default:  $-0.49 \mu\text{V}$ ; specific:  $-0.002 \mu\text{V}$ ;  $t(39) = -2.75, p = .009$ ]. However, the second interaction did not reveal any relevant effects.

In this early time interval, a P200 *gender congruity* effect for the default reflexive pronoun (他自己/*himself*) in the mid-anterior region was observed (congruent > incongruent) (see Figure 1 & Figure 4) and a P200 *reflexive pronoun gender specificity* effect (specific > default) emerged in the lateral anterior region (see Figure 2 & Figure 3).



**Figure 2. The averaged brain activities separately presented by reflexive pronoun gender specificity (default: 他自己/*himself* vs. specific: 她自己/*herself*) in the representative electrodes.**



**Figure 3. Topographic distributions for the reflexive pronoun gender specificity effects (她自己/herself minus 他自己/himself) respectively in the P200 and N400 time intervals.**

### 2.3.2 ERPs on the Reflexive Pronouns at the N400 Time Interval: 250 ms to 600 ms

The repeated measures ANOVA on the amplitude of the evoked activity showed the following pattern: For electrodes in the midline region, a main effect of *reflexive pronoun specificity* revealed that 她自己/herself was significantly more negative than 他自己/himself [default:  $-1.56 \mu\text{V}$ ; specific:  $-2.04 \mu\text{V}$ ;  $F(1, 39) = 7.33, p = .010$ ] and an interaction between *gender congruity* and *latitude* [ $F(2, 78) = 4.55, p = .028$ ] were found. However, this interaction did not reveal any relevant effects in the paired comparisons.

For the electrodes in the lateral scalp, the ANOVA showed two interactions. The first one showed that the *antecedent gender type* interacted with *latitude*, [ $F(2, 78) = 3.81, p = .035$ ]. However, the paired comparisons did not show any significant differences. The second one was an interaction among *reflexive pronoun specificity*, *latitude* and *lateral scalp* [ $F(2, 78) = 4.19, p = .019$ ]. The paired *t*-tests showed that 她自己/herself was significantly more negative than 他自己/himself in the left central region [default:  $-1.37 \mu\text{V}$ ; specific:  $-1.72 \mu\text{V}$ ;  $t(39) = 2.18, p = .035$ ] and no relevant effects were found in other lateral regions.

Considering the whole experimental design, no interaction between the factors of interest emerged in the N400 time interval. Only an independent N400 *reflexive pronoun specificity* main effect was found in which the specific reflexive pronoun (她

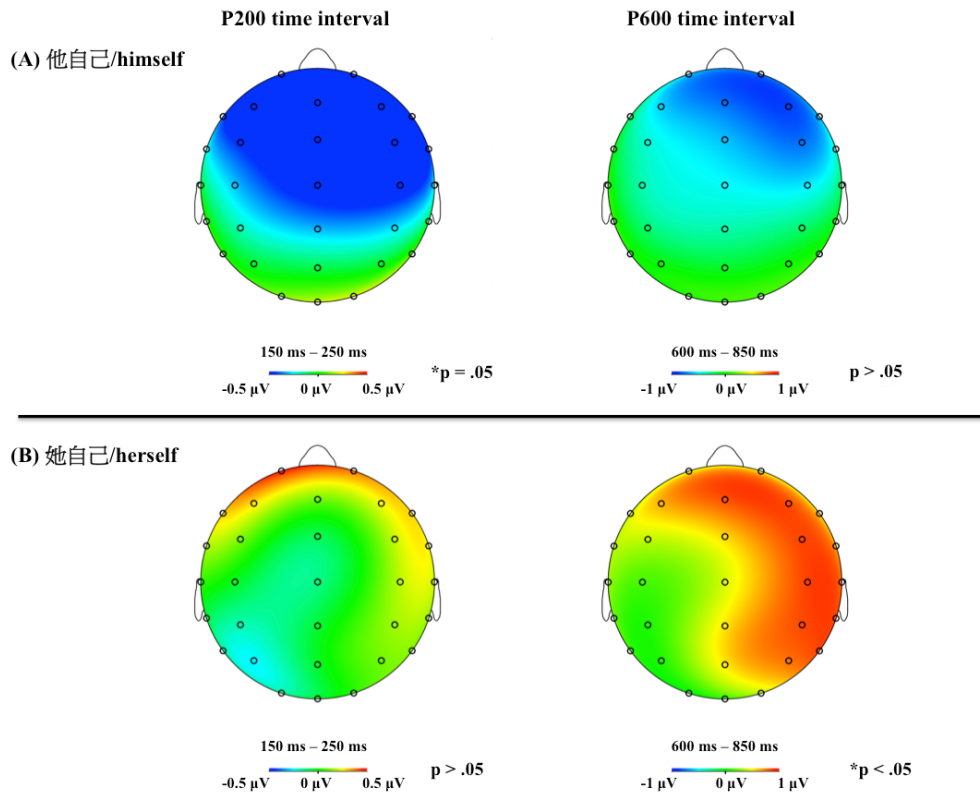
自己/*herself*) elicited more negative waveforms than the default (他自己/*himself*) (see Figure 2 & Figure 3).

### 2.3.3 ERPs on the Reflexive Pronouns at the P600 Time Interval: 600 ms to 800 ms

The repeated measures ANOVA for electrodes in the midline region revealed three interactions. First, *antecedent gender type* was found to interact with *gender congruity* [ $F(1, 39) = 4.53, p = .040$ ], but the paired *t*-tests did not show any significant effects. Second, the *reflexive pronoun specificity* interacted with *gender congruity* [ $F(1, 39) = 6.21, p = .017$ ]. The planned paired *t*-tests showed that the incongruent 她自己/*herself* elicited more positive amplitude as compared to the congruent one [congruent:  $-1.02 \mu\text{V}$ ; incongruent:  $-0.38 \mu\text{V}$ ;  $t(39) = -2.10, p = .042$ ], but no such difference was found for 他自己/*himself* [congruent:  $-0.22 \mu\text{V}$ ; incongruent:  $-0.61 \mu\text{V}$ ;  $t(39) = 1.26, p = .215$ ]. Third, an interaction among *reflexive pronoun specificity*, *gender congruity* and *latitude* [ $F(2, 78) = 5.08, p = .015$ ] emerged. The paired *t*-tests showed only a significant difference for incongruent 她自己/*herself* as compared to the congruent 她自己/*herself* in the anterior region [congruent:  $-1.77 \mu\text{V}$ ; incongruent:  $-0.85 \mu\text{V}$ ;  $t(39) = -2.39, p = .022$ ].

The repeated measures ANOVA for electrodes in the lateral scalp showed two interactions. The first one showed that the *reflexive pronoun specificity* interacted with *gender congruity* [ $F(1, 39) = 8.84, p = .005$ ]. The paired *t*-tests showed that the incongruent 她自己/*herself* elicited more positive amplitude as compared to the congruent one [congruent:  $-0.67 \mu\text{V}$ ; incongruent:  $-0.12 \mu\text{V}$ ;  $t(39) = -2.56, p = .015$ ], but no such effects emerged for 他自己/*himself* [congruent:  $-0.21 \mu\text{V}$ ; incongruent:  $-0.46 \mu\text{V}$ ;  $t(39) = 1.31, p = .198$ ]. The second interaction was among *reflexive pronoun specificity*, *gender congruity* and *lateral scalp* [ $F(1, 39) = 6.77, p = .013$ ]. The paired *t*-tests revealed that the incongruent 她自己/*herself* was more positive than the congruent 她自己/*herself* in the right hemisphere [congruent:  $-0.86 \mu\text{V}$ ; incongruent:  $-0.07 \mu\text{V}$ ;  $t(39) = -3.26, p = .002$ ].

The most relevant finding in this P600 time interval was the interaction between *reflexive pronoun gender specificity* and *gender congruity*. While the amplitude for incongruent 她自己/*herself* was more positive than that for congruent 她自己/*herself*, no such effects emerged for 他自己/*himself* (see Figure 1 & Figure 4).



**Figure 4.** Topographic distribution of the gender congruity effect (incongruent minus congruent) respectively for the two reflexive pronouns in the P200 and P600 time intervals. Panel (A) represents effects for the general reflexive pronoun, 他自己/*himself*, and panel (B) represents effects for the specific reflexive pronoun, 她自己/*herself*.

## 2.4 Discussion

The present study evaluates how the asymmetry of gender specificity encoded in characters for Chinese reflexive pronouns influences the processing of gender information during reflexive pronoun resolution in Chinese text. The results showed: First, two independent effects for 她自己/*herself*; a P200 effect (她自己/*herself* is more positive than 他自己/*himself*) and a N400 effect (她自己/*herself* is more negative than 他自己/*himself*), supporting the assumption of 他自己/*himself* as the default pronoun. Second, a P200 gender congruity effect for 他自己/*himself* (congruent is more positive than incongruent) and a P600 gender congruity effect for 她自己/*herself* (incongruent is more positive than congruent) also emerged. The dissociation between 他自己/*himself* and 她自己/*herself* in the two time intervals provides relevant evidence of the asymmetry of gender specificity for anaphor resolution and suggests the distinct time courses involved in the processing of gender information during reflexive pronoun resolution. Most importantly, such dissociation further clarifies that the default pronoun 他/*he* is perceived as equally applicable to both genders instead of being a male-biased pronoun because no gender congruity effects emerged in the semantic-related N400 or the semantic/integration-related P600 time intervals.

### 2.4.1 The Processing of Chinese Third Person Reflexive Pronouns

As predicted, the less familiar and less frequent reflexive pronoun, 她自己/*herself* elicited more positive amplitude in the P200 time interval and more negative amplitude in the N400 time interval. The results thus reveal that the semantic radicals are processed at an early perceptual phase during Chinese orthographic recognition and lexical access is faster and easier for the default pronoun 他自己/*himself*, as compared to the specific 她自己/*herself*.

A straightforward explanation for the P200 and N400 specificity results could be that these arise because of frequency effects, due to the large difference in the word frequencies between the two characters (see Table 1 & Figure 2) (Dambacher, Kliegl, Hofmann & Jacobs, 2006; Kutas & Federmeier, 2011). According to Dambacher and colleagues (2006), the P200 could index processing differences resulting from word frequency comparisons (low frequency words elicit more positive amplitude) and the “[...] *lexical access was presumably completed for high-frequency words while low-frequency words were still being processed*” (p. 96). Meanwhile, N400 word frequency effects were reported for low frequency words as compared to high ones (Van Petten & Kutas, 1990; Dambacher, Kliegl, Hofmann & Jacobs, 2006), even in the context of word repetition in sentences (Van Petten et al., 1991). Although word frequency could explain the P200 and N400 effects reported here, the mechanism of how the two reflexive pronouns are processed and the dissociation of the two reflexive pronouns in the P200 and P600 time intervals are not clarified by this explanation, taking into account the co-reference between a reflexive pronoun and its antecedent within the structure of anaphor resolution.

Instead, the attention-related feature analysis (P200) and semantic expectation/predictability (N400) viewpoints may well explain the processing mechanism and the dissociation of the two reflexive pronouns during anaphor resolution. The P200 gender specificity effect could be interpreted as an attention-related mapping cost of the high similarity of graphic form in the two reflexive pronouns during Chinese character recognition (他自己/*himself* vs. 她自己/*herself*) (Luck & Hillyard, 1994; Liu, Perfetti & Hart, 2003). Liu and colleagues (2003) manipulated graphic similarity between a prime, 凉/*liáng/cool*, and target, 惊/*jīng/startled*, in a pronunciation task and reported a P200 effect for the change in semantic radical. Following the attention-related feature analysis interpretation of Luck and Hillyard (1994), the authors concluded that the P200 effect is related to orthographic and phonological processing. This may possibly be the case in the 他自己/*himself*-她自己/*herself* contrast at the word level. As a default, the graphic form of 他自己/*himself* may be more familiar to participants and so make it easier to recognize/retrieve. In contrast, when 她自己/*herself* is encountered, participants need more effort to process the relatively less familiar graphic as compared to the default. This effect could be interpreted as basically driven by the semantic radicals encoded in the two reflexive pronouns.

Hung and Schumacher (2012; 2014) reported similar P200 effects in studies comparing the topicality (the amplitude for novel-topic was more positive than topic-shift, that was more positive than topic-continuity) and topic-worthiness (new topic was more positive than given topic) effects in Chinese discourse processing. The authors interpreted the P200 effect as reflecting early perceptual processing costs during discourse inference, taking into account the topicality or topic-worthiness. According to Hung and Schumacher (2014), the P200 “[...] *is likely to be a neural*

*response to the involvement of selective attention that facilitates perceptual processing of an item that fulfilled contextually-induced expectation”* (p. 43). Due to the fact that similar sentence structures were employed in the present experiment, the default reflexive pronoun may be considered as a given topicality expected after every antecedent mentioned in the previous sentence fragment. The 他自己/*himself* as a default may account for this context-induced expectation of discourse inference. It could facilitate the perceptual processing and result in a neural brain response to its antecedent because it already fulfills the expectation driven by the contextual information. The more positive amplitude observed for 她自己/*herself* would accordingly reflect the processing of an unexpected item detected by the brain during early perceptual processing and result in a P200 effect, thus in line with Hung and Schumacher (2012; 2014).

On the other hand, the fact that the N400 effect (她自己/*herself* is more negative than 他自己/*himself*) could reflect the differential semantic expectation/predictability between the default (referring to any human) and the specific (referring to females only) reflexive pronouns. Since 他自己/*himself* is applicable to every mentioned antecedent, less negative amplitude is expected, either due to its all-inclusive semantic meaning (Hagoort, Hald, Bastiaansen & Petersson, 2004; Lau, Phillips & Poeppel, 2008; Rabovsky & McRae, 2014) or its more accessible orthographic form (see Delong, K.A., Urbach & Kutas, 2005: the expected article “a” or “an” in English, cf. Kutas & Federmeier, 2011).

#### **2.4.2 Anaphor Resolution in a Language without Inflectional Morphological Gender Markings**

Taking into account the gender incongruity effects reported here, in terms of reflexive pronoun resolution, the present findings are consistent with Osterhout and colleagues (1997) in two respects. First, antecedent gender type (definitional as opposed to stereotypical gender) has no differential effect on the processing of gender information in anaphoric reflexive pronouns. Second, the processing cost for mismatches on gender emerges in the P600 time interval during reflexive pronoun resolution. According to Osterhout and colleagues (1997), although there is variability between definitional and stereotypical gender, as long as the gender information is activated (either male or female), the co-reference of this gender information during anaphor resolution should not vary by antecedent gender type. Thus, due to the syntactic constraints for definitional role nouns, and similar mismatch P600 effects found for both definitional and stereotypical role nouns, the authors concluded: gender information is “ [...] *encoded within grammar*” (p. 282) and results in syntactic processing. Following Osterhout and colleagues, Kreiner, Sturt and Garrod (2008, Experiment 1) further argued that the gender representation for such inference differed by noun types (i.e., definitional: lexical and stereotypical: pragmatic). Even so, Kreiner and colleagues (2008) reported no differences on the reflexive pronouns between sentences with definitional roles and those with stereotypical role nouns. Accordingly, our finding that no significant effects related to antecedent gender type emerged in any time interval very likely corresponds to what these authors described about the assignment of gender information regardless of antecedent noun type (definition or stereotypical) or how gender is represented by the antecedent (lexical or pragmatic). As long as the antecedent’s gender is assigned, the activated gender is taken for the following anaphor resolution. However, in the present study, the P600 gender congruity effect found for 她自己/*herself* cannot be interpreted as reflecting



syntactic processing, but is, rather, semantically-driven. As no morpho-syntactic gender agreement is required for Chinese anaphor resolution, the sentence remains grammatical when the character 他 is replaced by 她. Only the meaning-based semantic radicals can account for the co-reference difficulty between antecedent and anaphor. This result is in line with Xu and colleagues' (2013) proposal that the P600 effect in Chinese anaphor resolution reflects an integration difficulty based on semantic anomaly instead of reflecting processing difficulties with syntactic violations. This semantics-based interpretation of the P600 effect has already been offered by studies that did not involve any extra syntactic manipulations in their experiments (Münte et. al., 1998; Kaan, Harris, Gibson, & Holcomb, 2000; Kolk, Chwilla, van Herten & Oor, 2003; Kaan & Swaab, 2003; Kim & Osterhout, 2005; van Herten, Kolk, & Chwilla, 2005; Callahan, 2008; Van de Meerendonk et. al, 2009; Molinaro et. al., 2012). Thus, while findings in morphological gender languages (e.g. Osterhout, Bersick & McLaughlin, 1997; Kreiner, Sturt & Garrod, 2008; Kreiner, Mohr, Kessler & Garrod, 2009) have shown the importance of syntactic gender agreement between antecedents and pronouns, the case of Chinese seems to be different. Reflexive pronoun resolution in written Chinese needs to rely on the semantic information unequally encoded in the orthographic forms of the characters, given that one orthographic form codes for a generic gender while the other codes specifically for a feminine representation. The generic gender encoded by the default pronoun 他/*he* is confirmed here to be a gender-neutral pronoun (appropriate to both genders) in line with the rating study that words with the radical 人/*rén/human* are rated as neutral (Cherng, Chang & Chen, 2009). Thus, mechanisms of co-reference of gender information between antecedents and reflexive pronouns in this study are modulated by the information carried by the semantic radicals denoting the different gender specificities of the pronouns.

The new finding here related to anaphor resolution is the P200 gender congruity effect found for the default reflexive pronoun 他自己/*himself*. This mid-anteriorly distributed P200 effect (congruent conditions elicited more positive amplitude as compared to incongruent ones) is different from the P200 reflexive pronoun gender specificity effect observed in the lateralized electrodes (她自己/*herself* is more positive than 他自己/*himself*, discussed in section 2.4.1), because the former effect is related to the gender of antecedents at sentence level. Thus, it seems that the conceptual processing of gender (the former P200) and the perceptual processing of character recognition (the latter P200) are processed in a similar time interval but are independent and associated with distinct scalp regions. Consistent with the attention-related mapping cost and context-induced expectation accounts, because of the similar sentence structures in the experimental stimuli, participants could expect the appearance of reflexive pronouns after encountering a proper name or a stereotypical role noun. If the antecedent is female, it is possible that participants are prepared for both reflexive pronouns, as both are applicable to a female antecedent. During the time interval in which the default (他自己/*himself*) is recognized, the evaluation of gender congruity is also easily accomplished because of its wide range of applicability. If the antecedent is male however, within this experimental context (because of the number of incongruent items of this nature presented), both the perceptual and conceptual processing systems might require more cognitive resources for encountering either 他自己/*himself* or 她自己/*herself*. The amplitude difference between the two contrasts (himself-congruent is more positive than himself-incongruent) might reflect such attention-related cost, related to prior contextual

information (i.e., antecedent's gender here) deployed for processing the reflexive pronoun (Luck & Hillyard, 1994; Blanchet, Gagnon & Bastien, 2007). In this case, the P200 is also sensitive to contextual information. On the other hand, when the specific reflexive (她自己/*herself*) is encountered, its recognition and retrieval are more complicated (as discussed in 2.4.1) as compared to the default. It is possible that the word recognition is accomplished in the P200 time interval (i.e., the P200 effect for 她自己/*herself*) and the evaluation of gender congruity is delayed and resolved in a later time interval (i.e., the P600 effect).

Taking the main findings together, it is clear that 他自己/*himself* as a default with neutral gender is a critical feature in the perceptual and conceptual processing of gender information during Chinese reflexive pronoun resolution. Both types of processing rely on the semantic radicals encoded in the characters, suggesting the essential importance of the gender-based radicals to decoding of gender specificity. In terms of gender congruity effects, only when there is a clear mismatch (i.e., male/male-biased antecedent followed by the specific female reflexive), do our results support a two-stage model of anaphor resolution (Garrod & Sanford, 1994). As a whole, given the more familiar orthographic form (perceptual) and applicability to both genders (conceptual) of 他自己/*himself*, this pronoun may serve as a baseline during anaphor resolution. In addition, the ERP pattern for Chinese reflexive pronoun resolution confirms distinct time courses of processing for the two reflexive pronouns. While the default, 他自己/*himself*, is processed mainly at the early perceptual stage of character recognition and gender evaluation, the processing of the specific reflexive pronoun, 她自己/*herself*, lasts from the early perceptual stage (bonding stage, possibly including the N400 time interval) to the late integration stage (resolution stage). It is possible that there is more than one way to resolve anaphors. One is a two-stage model when the gender of an anaphor is specific and the mismatch is definite. The other is a one-stage processing model in which a default anaphor is eligible for every mentioned antecedent, as is the case of 他自己/*himself* reported in the present study. This does not mean that these results are specific to languages with default and specific distinctions on pronouns, as in Chinese. Instead, in addition to the well-established two-stage model for anaphor resolution (when pronouns have a specific gender), the one-stage model extends the description of anaphor resolution to when a pronoun can be used for both genders (in other words, a genderless pronoun). According to Siewierska (2013), genderless pronouns are used in 67% (254 out of 378) of world languages (such as Finnish, Turkish, Thai, Indonesian, Vietnamese or Maori ... etc.). The results observed here may thus be relevant to the 67% languages with genderless pronouns.

## 2.5 Conclusion

The non-symmetrical gender specificity of the Chinese characters for third person reflexive pronouns was studied during anaphor resolution. Independent P200 and N400 gender specificity effects confirmed processing differences resulting from the different gender specificity of reflexive pronouns (encoded in their semantic radicals) and also suggested the functional role of 他自己/*himself* as a default during anaphor resolution. During reflexive pronoun resolution, the two types of gender specificity interact with gender congruity respectively in the P200 (他自己/*himself*) and P600 (她自己/*herself*) time intervals. These results provide further evidence in support of the two-stage model of anaphor resolution only when there is an

unambiguous mismatch between the antecedent and anaphor. The ERP patterns of the two reflexive pronouns also highlight the distinct time courses of anaphor resolution resulting from the two types of gender specificity. Overall, the findings in the present study demonstrate the importance of taking into account the asymmetry of gender specificity in Chinese third person reflexive pronouns, as well as confirming 他自己 /*himself* as the default applicable to both genders.

## 2.6 Problems Unsolved and the Following Experiment

The present experiment has analyzed anaphor resolution studies by manipulating the antecedent of stereotypical role nouns in a genderless language. The present findings complete anaphor resolution studies that are now available in all the three language groups (see Table 7 in Chapter 5 for details).

A closer look at Table 7, it appears that during anaphor resolution, the processing of the gender feature for antecedent of definitional role nouns is consistent with the findings reported in other language groups (Mandarin Chinese – semantics-based; English – syntactic; German – syntactic but possibly interacting with semantics). Regarding antecedent of stereotypical role nouns, language-gender marking consistency determines the same effects in Mandarin Chinese and English, but not in German (i.e., semantics-based, Irmen, Holt & Weisbrod, 2010<sup>8</sup>). Why? Based on Irmen and colleagues (2010), the effect in German is due to the gender mismatch between the referential gender (stereotypical) of the antecedent (e.g., *viele Informatiker/many computer scientists*, male-biased) and the lexical gender of the anaphoric noun phrase (e.g., *diese Frauen/these women*). A simple explanation for this discrepancy could be that as only semantic information is manipulated in this study, the processing correlate reflects semantics-based processing. Yet, it should be noted here that the referential gender is opposite to the morpho-syntactic gender presented in the antecedent (i.e., a stereotypical noun phrase). In grammatical gender languages, gender agreement is morphosyntactically required. Mismatches on grammatical gender in non-animate noun phrases could thus result in syntactic processing difficulties (LAN plus P600, Gunter, Friederici & Schriefers, 2000; Barber & Carreiras, 2005; Molinaro, Vespignani & Job, 2008; O'Rourke & Van Petten, 2011; Caffarra, Siyanova-Chanturia, Pesciarelli, Vespignani & Cacciari, 2015). How could this referential gender be processed beyond morpho-syntactic gender presented in German noun phrase?

This kind of gender-inconsistent activation has been reported previously in Carreiras, Garnham, Oakhill and Cain (1996). The authors proposed that the default stereotypical gender is activated as long as the related information is available regardless of the morpho-syntactic gender morphology present in the role noun (immediacy hypothesis). The authors compared participants' self-paced reading time during pronoun resolution and found that while Spanish speakers' reading time was slowed down when they encountered stereotypical gender inconsistent role nouns in the antecedent position, English speakers' reading time was slowed down when they read stereotypical gender inconsistent pronoun at the anaphora position (Carreiras,

<sup>8</sup> According to Irmen, Holt & Weisbrod (2010), “[...] the anaphoric NPs are formally correct and furthermore permit a gender differentiation in contexts with plural antecedents. Anaphor resolution here depends on the fit between the referential gender of the antecedent and the referential gender of the anaphor. Anaphor resolution here does not depend on linguistic form or any syntactic constraint.” (p. 135). “[...] NPs were chosen as anaphoric references in order to constrain anaphor resolution only semantically and not syntactically” (p. 141).

Garnham, Oakhill & Cain, 1996). This means that the activation of stereotypical gender has occurred already at the antecedent position and not necessarily at the pronoun position (see section 1.2.3, paragraph with more details). Again, the same question emerges, how does the default stereotypical get activated without the presence (or indication) of morpho-syntactic gender?

Based on Carreiras and colleagues (1996) and Irmen and colleagues (2010), it seems that morpho-syntactic gender morpheme has a limited role in the activation of the default stereotypical gender. How could this happen in grammatical gender languages, in which the morphosyntactic gender is so prominent? Since grammatical gender was not manipulated in the two studies here described, we do not know how the processing would be in the case that gender agreement is additionally manipulated. Thus, behind stereotypical processing during anaphor resolution, how stereotypical knowledge interacts with the morpho-syntactic gender morphemes presented in stereotypical role nouns in grammatical gender languages turns to be the focus now. How default stereotypical gender is processed if grammatical gender is also manipulated within the same noun phrase (determiner + role noun). This would put more light on the immediate hypothesis proposed by Carreiras, Garnham, Oakhill and Cain (1996).

In the following experiment, we decided to directly measure the online neurocognitive processing of stereotypical role nouns in a grammatical gender language (Spanish here). Stereotypical gender and grammatical gender will be simultaneously manipulated within the same noun phrase in sentences. The ERP mismatch paradigm will also be used to test the immediacy hypothesis.

## **CHAPTER 3**

### **Experiment II: The Online Processing of Stereotypical Gender in a Grammatical Gender Language**

### 3.1 Introduction<sup>9</sup>

Many words are associated to gender-oriented stereotypes, i.e., beliefs concerning the gender properties of certain social groups (English: Banaji & Hardin, 1996; Kreiner et al., 2008; Spanish: (Carreiras, Garnham, Oakhill & Cain, 1996; German: Irmen & Roßberg, 2004). Stereotypical knowledge is automatically activated (and difficult to inhibit) for words referring to people (*nurse, doctor*; Oakhill et al., 2005), but also for object entities (*bikini, cigar*; Garnham et al., 2002) and adjectives (*aggressive, nurturing*; White et al., 2009). Many studies have shown that the activation of stereotypical knowledge is immediate and not based on conscious inferences (Carreiras et al., 1996; Garnham et al., 2002). Nonetheless, the influence that stereotypical cues exert on language processing is still a matter of debate. In fact, while a high number of studies have focused on the interaction between other sources of semantic knowledge and syntactic information during language processing (e.g., Osterhout & Nicol, 1999), it is unknown how simultaneous stereotypical and syntactic cues modulate language comprehension processes. Interestingly, the distinction between stereotypes and other types of semantic knowledge is mainly motivated by the fact that the brain processes information about categories of objects and living things (i.e., semantic knowledge) differently from information about categories of people (i.e., stereotypes; Contreras et al., 2012, and references therein). In the present study, we employed event-related potentials (ERPs) to evaluate the processing of stereotypes and of morphosyntactic cues.

Previous studies on stereotype processing during sentence comprehension have focused on anaphoric constructions. Osterhout and colleagues (1997) reported qualitatively similar ERP effects for pronouns mismatching with either the gender definition (*mother, father*) or the gender stereotype (*nurse, doctor*) of the previous antecedent. They reported increased positive amplitude shifts starting after 600 ms in the posterior scalp regions that they identified as P600s (similar results in Canal et al., 2015; Su et al., 2016). While some authors have claimed that the P600 reflects syntactic processing (e.g., Kim & Osterhout, 2005), recent views suggest that it can reflect more general language integration (e.g., Brouwer et al., 2012), repair and reanalysis (e.g., Friederici, 2011) or more general conflict monitoring (e.g., van de Meerendonk et al., 2010; for a review Kuperberg, 2007). Independently of the functional interpretation of the P600 effect, the similar ERP correlates observed for stereotypical and definitional gender led Osterhout and colleagues (1997) to conclude that stereotypical cues syntactically constrain pronoun resolution (see similar claims by Esaulova et al., 2014).

Osterhout et al.'s (1997) conclusions were mainly driven by the fact that no semantic-related ERP effect was observed for the stereotypical condition, in which world-knowledge inferences (known to elicit semantic ERP correlates such as the N400, Hagoort et al., 2004) could potentially have driven pronoun resolution processes. A number of studies have indeed reported stereotype-related modulations of the N400 ERP component triggered by isolated role nouns in semantic priming paradigms (e.g., Siyanova-Chanturia et al., 2012). The N400 is a negative-amplitude deflection around 400ms, more evident in right posterior brain regions (Kutas & Hillyard, 1984). It is sensitive to a large number of lexical-semantic parameters and has also been associated to stereotype congruity. White and colleagues (2009), for

<sup>9</sup> The present experiment is published in *Brain and Language* and the reference is as follows: Molinaro, N., Su, J.-J., and Carreiras, M. (2016). Stereotypes override grammar: Social knowledge in sentence comprehension. *Brain and Language*, 155/156, 36-43.

example, reported a stereotype-related negativity and interpreted this as a N400 effect even though the distribution of the effect was evident across the whole scalp, including anterior electrodes.

Potential explanations for the lack of the N400 effect in Osterhout et al. (1997) rely on the fact that the authors did not estimate stereotype processing by focusing on the ERPs time-locked to the role noun, but to a distant pronoun. Given that previous behavioral studies have shown the immediacy with which stereotypical knowledge is activated upon reading a word (Carreiras et al., 1996; Garnham et al., 2002), it is relevant to evaluate the ERP correlates time-locked to a stereotypical target word. As an example, in an eye-tracking experiment, Kreiner and colleagues (2008) observed processing differences between definitional and stereotypical gender in cataphoric (but not in anaphoric) constructions in which pronouns preceded the stereotypical nouns where the experimental effects were measured.

In contrast to previous studies that separately compared stereotypical and definitional anaphoric mismatches, we here analyzed the ERP correlates triggered by the reading of stereotypically-biased role nouns, while manipulating the gender of the target word (identified by the noun ending) and the syntactic context preceding the target word (the gender marked determiner). We followed the rationale of ERP studies that evaluate the interaction between syntactic and semantic mismatches in a factorial design. The literature focusing on the interaction between syntax and semantics has observed dissociable ERP correlates for semantic and syntactic violations. The double mismatch condition has either confirmed the independence and autonomy between semantic and syntactic processing, showing additive effects (Gunter et al., 2000; Osterhout & Nicol, 1999) or has provided evidence for the vulnerability of the semantic processor to syntactic cues, but not the opposite (Hagoort, 2003). Martin-Loeches et al. (2006) proposed the relative prevalence of semantics over syntax in Spanish (the language investigated in the present experiment). However, they reported qualitatively different effects for syntactic and semantic errors and an additive effect for the double anomaly when considering the pre-stimulus ERP baseline (as Osterhout & Nicol, 1999).

Brain sensitivity to the stereotypicality of the target word is studied in the present experiment by manipulating the role noun ending (congruent: *miner-os*, male miners; incongruent: *miner-as*, female miners). The syntactic context is manipulated by taking advantage of the gender markedness of the Spanish article, so that it could agree with a following role noun (...*los mineros*..., the[+m] male miners) or not (\*...*las mineros*..., the[+f] male miners). The double anomaly condition (\*...*los mineras*..., the[+m] female miners) will provide additional evidence about the interaction between stereotypical and syntactic cues during language comprehension (see Table 3).

**Table 3. Example of the Experimental Sentences**

Condition	Example sentences
Control ( <i>stereotypically congruent/syntactic agreement</i> )	Ayer, <b>los mineros</b> fueron a una cena para celebrar el fin de la asamblea Yesterday, the <sub>[+M]</sub> ( <i>male</i> ) miners went to a dinner for the celebration of the end of an assembly
Stereotypically incongruent ( <i>and syntactic agreement</i> )	Ayer, <b>las mineras</b> fueron a una cena para celebrar el fin de la asamblea Yesterday, the <sub>[+F]</sub> ( <i>female</i> ) miners went...
Syntactic violation ( <i>and stereotypically congruent</i> )	*Ayer, <b>las mineros</b> fueron a una cena para celebrar el fin de la asamblea *Yesterday, the <sub>[+F]</sub> ( <i>male</i> ) miners went...
Double anomaly ( <i>stereotypically incongruent/syntactic violation</i> )	*Ayer, <b>los mineras</b> fueron a una cena para celebrar el fin de la asamblea. *Yesterday, the <sub>[+M]</sub> ( <i>female</i> ) miners went...

We contrasted two possible outcomes for the present experiment. On the one hand, stereotypical knowledge could be processed in the same way as other sources of semantic knowledge, so that we should observe a relative independence between the ERP effects elicited by grammatical and stereotypical anomalies (Gunter et al., 2000; Martin-Loeches et al., 2006; Osterhout & Nicol, 1999; but also see Hagoort, 2003). We would therefore expect different ERP correlates for the stereotypicality and the syntactic manipulations, i.e., an increased N400-like effect for stereotypes (White et al., 2009), and for grammatical anomalies a P600 that could be preceded by a Left Anterior scalp distributed Negativity (LAN) around 400 ms, an ERP correlate typically observed for morphosyntactic anomalies in morphology-rich languages (Barber & Carreiras, 2005; Molinaro et al., 2008; 2011a; 2011b; 2015). This last prediction is motivated by previous studies which employed a similar linguistic construction (determiner-noun gender agreement error involving animate nouns, e.g., Sabouring & Stowe, 2008) and by previous research on stereotypes involving gender agreement mismatches (Canal et al., 2015; Osterhout et al., 1997; Su et al., 2016). In the double anomaly condition, additive ERP effects could be observed (Gunter et al., 2000; Martin-Loeches et al., 2006; Osterhout & Nicol, 1999).

On the other hand, if we take to the extreme the “immediacy hypothesis” discussed by some authors (Carreiras et al., 1996; Garnham et al., 2002), stereotypical knowledge (activated by the word stem *miner-*, male stereotyped) could possibly have more weight than other forms of linguistic cues. In this case, we would expect that any incongruence with the stereotypical gender of the role noun (i.e., either morphological, the noun ending *-as*, or syntactic, the previous gender marked determiner, *las*) would trigger a similar brain reaction. Plausibly, this effect would be the N400-like effect observed for isolated word stereotypical processing (White et al., 2009). Under this assumption, no syntactic ERP correlates should be observed (i.e., LAN and P600) even in sentences involving grammatical errors.



## **3.2 Materials and Methods**

### **3.2.1 Participants**

Twenty-four native Spanish speakers (12 males, mean age: 23.21, range 18-30) took part in the ERP study and were paid for their participation. They were healthy, right-handed, with normal or corrected-to-normal vision. None of them reported prior history of neurological disorder.

### **3.2.2 Materials**

The stereotypical role nouns used in this study were selected from questionnaire containing 328 Spanish role nouns. We used plural nouns to avoid potential differences in the number of letters between feminine and masculine stimuli. Sixty-four native Spanish speakers (32 males, aged from 18 to 38, mean age: 21.83) were paid for rating this questionnaire. Participants were instructed to judge on an 11-point Likert scale (from 0% to 100% with 10% intervals) the percentage of males/females composing a category based on their immediate intuition. The 40 most male-biased, and the 40 most female-biased role nouns were selected as target words (Appendix C: 1-40 were male-based and 41-80 female-biased). The eighty noun phrases (determiner + role noun) were inserted in 80 non-constraining sentences. We manipulated in a factorial design the Stereotype Congruity of each role noun (StereoCong: congruent vs. incongruent) and the Gender Agreement of noun phrase (GendAgree: agreement vs. disagreement). Across items, sentence structure was similar (Table 3) and the target role nouns were located from the third to the fourteenth sentence position but never in the sentence final position.

Stimuli were balanced across participants employing four lists with equal number (40) of trials in each condition, so that each participant did not see the same item in different conditions. Additional filler sentences were employed to balance the number of correct/incorrect sentences, for a total of 334 sentences per list.

### **3.2.3 Procedure**

Individual words were presented in white letters on a dark grey background. After a 500 ms centered fixation point followed by a 200 ms blank screen, each word was presented for 300 ms and 200 ms blank followed. The inter-sentence time interval varied from 1700 ms to 3000 ms.

During the experiment, participants sat in a sound-attenuated cabin and were instructed to read each sentence silently and carefully. Their task was to answer yes/no comprehension questions by button pressing. Comprehension questions were introduced after 30% of the sentences. Questions referred to the content of each sentence and never to the gender-related information of our interest. Twelve practice trials preceded the beginning of the experiment.

### **3.2.4 Data Recording and Analysis**

EEG data were recorded by the BrainAmp system. The 27 active electrodes were positioned on a cap according to the 10-20 system and the impedance was kept below 5 $\Omega$  for all electrodes. The sampling rate (A/D) was 1000 Hz. The on-line reference electrode was the left mastoid (M1). Signals were amplified with a bandpass of 0.01-

250 Hz. HEOGs were placed at the outer canthi of the eyes and VEOGs were placed above and below the left eye.

Brain Vision Analyzer was used to perform off-line signal processing analyses. EEG data were off-line re-referenced to the average activity of M1-M2. Signal was bandpass filtered between 0.3 and 30 Hz. Epochs of interest lasted from –250 ms before the onset of the target words to 1000 ms post-stimulus onset (–250 to 0 ms baseline corrected). Artifact activity was excluded based on visual inspection of each trial. As a result, 10% of the trials were removed (there were no differences in rejection rates among experimental conditions).

Statistical analyses (R software) focused on the mean voltage at each electrode within time intervals of interest. We focused on the time intervals corresponding to the N400/LAN (300–600 ms) and the P600 (600–850 ms). Further analyses were, however, carried out on the time intervals corresponding to early visual components (C1: 0–80 ms; N1: 80–150 ms; P2: 150–250 ms) to evaluate the presence of possible perceptual effects.

Repeated measures-ANOVA was performed considering the two experimental factors (StereoCong and GendAgree) and two topographical factors: Hemisphere (three levels: left, medial, right) and Latitude (three levels: anterior, central, posterior). This pattern considered nine groups of three electrodes (see groups in the upper panel of Figure 5). We report in the Results section significant (Greenhouse-Geisser corrected) effects involving the experimental conditions. Possible interactions between the two experimental factors were resolved comparing the different anomalous conditions to the control condition (FDR corrected pairwise *t*-tests).

### **3.3 Results**

#### **3.3.1 Comprehension Questions**

Participants responded accurately to the end-of-sentence questions with an average accuracy of 86% (range: 80-90%).

#### **3.3.2 ERPs Time-locked to Role-nouns**

No statistically reliable effects emerged from the ANOVA when considering the early time intervals (< 250 ms). In

Figure 5 it is possible to appreciate the increased negative effect peaking around 400 ms for all the incongruent conditions compared to the correct control sentences.

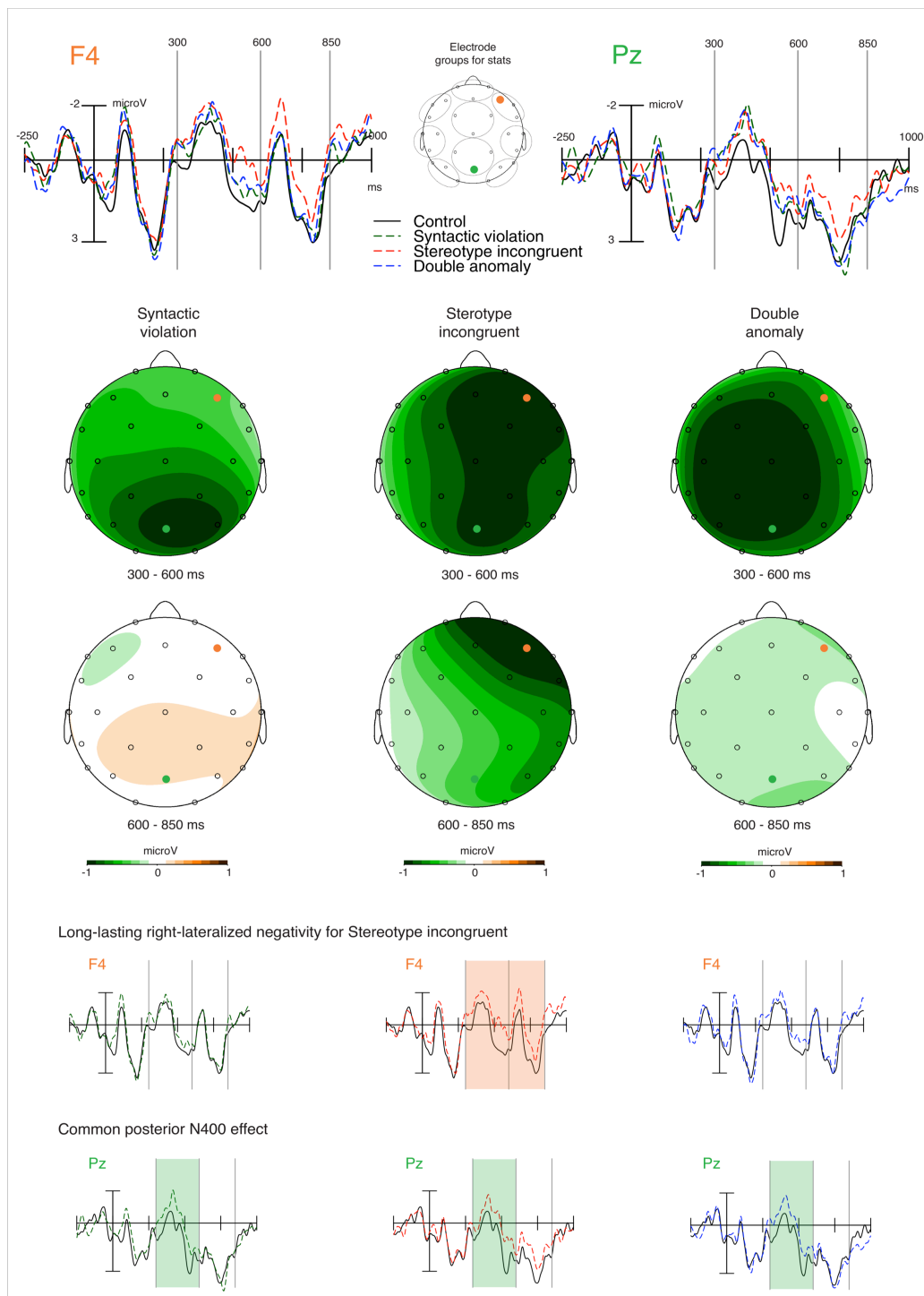
##### **3.3.2.1 The 300-600 ms Interval**

The interaction between StereoCong and GendAgree was statistically significant [ $F(1,23) = 8.01, p < 0.01, \eta^2 = 0.01$ ]. Post-hoc comparisons (considering *all* electrodes) showed that the stereotypically incongruent [mean difference (m.d.):  $-0.77 \mu\text{V}; t(23) = 3.15, pFDR < 0.01$ ], the syntactic violation [m.d.:  $-0.75 \mu\text{V}; t(23) = 3.45, pFDR < 0.01$ ] and the double anomaly [m.d.:  $-0.83 \mu\text{V}; t(23) = 2.79, pFDR < 0.05$ ] all differed from the control condition. In addition, a triple interaction between StereoCong, GendAgree and Hemisphere emerged [ $F(2,46) = 3.27, p < 0.05, \eta^2 = 0.01$ ]. Independent (three-way) ANOVAs excluding the Hemisphere factor for the left, medial and right electrode groups revealed that the interaction between StereoCong and GendAgree was stronger in the right hemisphere [right:  $F(1,23) = 11.22, p < 0.01,$

$\eta^2 = 0.03$ ; left:  $F(1,23) = 3.23, p < 0.1, \eta^2 < 0.01$ ; medial:  $F(1,23) = 5.67, p < 0.05, \eta^2 < 0.01$ ]. In the right electrodes a significant difference emerged between the stereotypically incongruent and the control condition [m.d.:  $-0.82 \mu\text{V}; t(23) = 2.98, pFDR < 0.01$ ; m.d. for syntactic violation:  $-0.46 \mu\text{V}$ , *not significant (n.s.)*; m.d. for double anomaly:  $-0.55 \mu\text{V}$ , *n.s.*].

### 3.3.2.2 The 600-850 ms Interval

The significant interaction between StereoCong and GendAgree [ $F(1,23) = 4.94, p < 0.05, \eta^2 = 0.01$ ] was driven by the sustained negative effect carrying on from the earlier time window [m.d.:  $-0.71 \mu\text{V}$ ] between the stereotypically incongruent condition and the control condition [ $t(23) = 2.58, pFDR < 0.05$ ; m.d. for syntactic violation:  $-0.15 \mu\text{V}$ , *n.s.*; m.d. for double anomaly:  $-0.25 \mu\text{V}$ , *n.s.*]. The triple interaction including Hemisphere emerged also in this time interval [ $F(2,46) = 9.21, p < 0.05, \eta^2 = 0.01$ ]. Separate ANOVAs in the three levels of the factor Hemisphere showed that the interaction between StereoCong and GendAgree was significant in the right electrodes only [right:  $F(1,23) = 11.99, p < 0.01, \eta^2 = 0.03$ ; left:  $F(1,23) = 0.59, n.s.$ ; medial:  $F(1,23) = 2.24, n.s.$ ]. The right electrodes interaction was driven by the stereotypicality contrast [m.d.:  $-0.96 \mu\text{V}; t(23) = 3.68, pFDR < 0.01$ ; ; m.d. for syntactic violation:  $-0.23 \mu\text{V}$ , *n.s.*; m.d. for double anomaly:  $-0.08 \mu\text{V}$ , *n.s.*].



**Figure 5. ERPs time-locked to the presentation of the target stereotypical noun. In the upper panel the four conditions are plotted in the same graph for two relevant electrodes. In the intermediate panels the topographical scalp distributions of each anomalous condition compared to the control in the 300–600 ms and 600–850 ms time intervals are plotted. In the lower panel we highlight the relevant effects of interest.**

### 3.4 Discussion

Two main findings emerge from the present study: first, the stereotypicality contrast elicited a long-lasting negativity starting in the N400 time interval and extending to right *anterior* scalp regions; second, we did not observe ERP correlates of syntactic processing for the comparisons involving gender agreement manipulations, but increased negative effects in posterior scalp electrodes, resembling a classic N400.

The scalp distribution of the stereotypicality effect (but also its long-lasting duration) involving additional right frontal regions is different from typical N400 lexical effects that are observed only in the posterior scalp regions. This suggests that stereotypical knowledge is represented differently from other forms of semantic knowledge. Stereotypes refer to knowledge about social groups, which seems to be emotionally more relevant to us compared to knowledge about non-social entities (Norris et al., 2004). Contreras and colleagues (2012) reported increased activation for brain regions involved in social cognition for social stereotypes, while knowledge about non-social entities activated regions associated with general semantic knowledge. More specifically, they propose that social stereotype processing depends on our ability to represent the mental state of the members of a group. Interestingly, this ability involves anterior frontal brain regions whose activity at the scalp level emerges as an increased negativity in the right lateralized electrodes (Sabbagh et al., 2004). This right anterior effect could overlap with the more posterior effect, more recognizable as a typical N400 effect, and give rise to the widespread and long-lasting effect observed for the stereotype contrast (Figure 5).

It could be argued, however, that this effect does not necessarily reflect stereotype processing but lexical frequencies of the targets. The stereotypically incongruent items that are less frequent compared to the congruent ones could have triggered the N400 effect. To test this claim, we separately considered masculine and feminine stereotypes. There is a general tendency to refer to role nouns in their masculine version and this interacts with stereotypicality. Indeed, masculine-congruent stereotypes were much more frequent than the incongruent counterparts (congruent: logFreq: 1.01; incongruent: logFreq: 0.45;  $t(39) = 4.10$ ,  $p < 0.01$ ), and masculine incongruent ones were numerically, but not statistically, more frequent than feminine congruent ones (congruent: logFreq: 0.52; incongruent: logFreq: 0.62;  $t(39) = -0.81$ , *n.s.*). Even so, the magnitude of the negative stereotypicality effect across all electrodes was similar in the two cases ( $-0.68 \mu\text{V}$  for the first and  $-0.86 \mu\text{V}$  for the second contrast). This shows that the amplitude of the negative stereotypicality effect was not sensitive to lexical frequency parameters. This difference is also informative of the relative asymmetry in the male-biased rates: female role nouns were less stereotypically female-biased compared to the male-biased role-nouns, which showed more extreme values (Appendix C). Despite this asymmetry, the stereotypicality effect was numerically larger for female role nouns.

Interestingly, these data also suggest that stereotypical knowledge strongly influences language processing functions. In fact, the similar posterior N400 effect across the three conditions emphasizes the pivotal role played by the stereotypes in interpreting the incongruities. In the three experimental contrasts, stereotypical knowledge of the role noun is activated independently of any morphological cue. If any such cue (either the previous determiner or the noun suffix or both) contrasts with the gender stereotype, the neurocognitive system reacts to the anomaly by anchoring on the stereotype. While this negative ERP effect was expected for the stereotypical

contrast (White et al., 2009; even if not often reported for sentence processing), it was unexpected for the anomalies involving morphosyntactic manipulations (associated to LAN-P600; Molinaro et al., 2011a; Martin-Loeches et al., 2006; see Barber et al., 2004, for agreement involving biological gender). Thus, the strong attracting force of the role noun's gender stereotype overshadows the typical ERP correlates of morphosyntactic violations such as gender agreement in Spanish (LAN-P600 biphasic response) and thus explains their absence in the present experiment.

A possible factor that could have triggered the N400 also for syntactic errors (differently from previous studies) is the presence of a large number of role nouns in whole experimental set of stimuli. It is not easy to disentangle whether the N400 for syntactic errors is inherent to the processing of each target role noun or is due to the context of the experiment, with a large number of role nouns. Future research should address this issue. However, it is worth mentioning that previous studies on gender stereotypes (in which role nouns abounded) consistently reported P600 effects (Canal et al., 2015; Osterhout et al., 1997; Su et al., 2016) and no N400s. The present study differs from previous research by focusing on a full factorial design, on a different syntactic construction and on a morphology-rich language like Spanish.

In sum, the present findings differ from previous studies that have investigated sensitivity to semantic and syntactic cues during morphosyntactic processing. These studies (Gunter et al., 2000; Hagoort, 2003; Martin-Loeches et al., 2006; Osterhout & Nicol, 1999) reported reliable sensitivity to the structural language cues independently of the interaction with different forms of semantic knowledge. In a similar vein, Van Berkum et al. (2008) reported N400 effects for semantic anomalies and speaker inconsistencies (stereotypically male utterances produced by a female speaker, see also Lattner & Friederici, 2003, for similar evidence involving P600). Despite the different experimental paradigm from the one used in the present study, the comparison between the two studies highlights the strong stereotypical force triggered by role nouns during reading.

In contrast, in the present study, stereotypical knowledge reduced the influence of syntactic cues (grammatical gender) in language comprehension processing. This unexpected result supports proposals indicating that social stereotypes should not be considered a typical form of semantic knowledge (Contreras et al., 2012). Future research could possibly examine possible modulating factors of the present stereotype effect, such as previous discourse context (Nieuwland & Van Berkum, 2006) or subjective variables such as empathy (Van den Brink et al., 2012) or social class (Varnun et al., 2012).

### ***3.5 Problems Unsolved and the Following Experiment***

The research question here shifts from how gender information is processed during anaphor resolution (Experiment I) to how stereotypical knowledge interacts with grammatical gender on the role nouns in grammatical gender languages. In a grammatical gender language, like Spanish, morpho-syntactic gender is expected to interfere during online processing. However, opposite to the prediction, stereotypical gender was found to dominate conflict resolution. Such results not only go against the general understanding of the leading role of morphosyntax in Spanish processing (in which P600 effects are reported, Barber & Carreiras, 2005; Molinaro et al., 2008; 2011a; 2011b; 2015), but also confirm the special status of stereotypical knowledge that is different from general semantic knowledge. Thus, compared with the linguistic

scenario of anaphoric pronoun resolution, it seems that processing of stereotypical gender is more visible on the role nouns rather than the pronouns.

In addition, the present findings support the immediacy hypothesis (a default stereotypical gender is activated regardless of inconsistent morpho-syntactic gender morphemes, Carreiras, Garnham, Oakhill & Cain, 1996). This may imply that morpho-syntactic gender morphemes (i.e., *-o* or *-a*) presented in stereotypical role nouns in grammatical gender languages merely function as a formal cue (grammatical and/or stereotypical) for participants to access the default stereotypical gender. Processing of stereotypical noun phrases is after all dominated by the unique semantic feature of stereotypical knowledge. If this is the case, can we generalize the immediacy hypothesis to genderless languages where morpho-syntactic gender morphemes are totally absent? Specifically, in a genderless language, such as Mandarin Chinese, the default stereotypical gender of a role noun is perceived via covert category<sup>10</sup> instead of relying on any formal gender-related marking (e.g., 矿工/*kuàng gōng/miner*, male-biased). That is, each generic role noun has a default gender and it is consequently recognized. Gender markers are independent words (e.g., 男/*male* or 女/*female*) required to appear before (i.e., to the left of) a generic role noun if one's gender is opposite to the default (e.g., '女' 矿工/*nǚ kuàng gōng/'female' miner*). No change is made on the generic role noun itself and no morpho-syntactic gender agreement is required. Since the two gender marking systems in Spanish and Mandarin Chinese are totally different from each other, would stereotypical knowledge keep its priority during the processing of stereotypical role nouns? To investigate this issue a similar experiment was carried out in Mandarin Chinese, a genderless language.

<sup>10</sup> Covert category depends on the conventional gender that is accepted through common practice in the society (Farris, 1988). To some extent, it is similar to the default stereotypical gender discussed in the text.

## **CHAPTER 4**

### **Experiment III: The Online Processing of Stereotypical Gender in a Genderless Language**



## 4.1 Introduction

The activation of stereotypical gender information is triggered immediately, as soon as the linguistic cues are accessible for stereotype activation and/or for re-activation of such stereotypes during pronoun resolution (Carreiras, Garnham, Oakhill & Cain, 1996; Osterhout, Bersick & McLaughlin, 1997). The “immediacy hypothesis” was outlined by Carreiras and colleagues (1996): stereotypical gender (that is acquired from society) is immediately activated as soon as participants perceive a stereotyped role noun, even when the grammatical gender expressed by the text does not match that information. This statement represents the strong view that social knowledge (i.e., stereotypical gender) drives sentence comprehension when contrasted with syntax (i.e., grammatical/syntactic gender). By employing the ERP mismatch paradigm, Molinaro, Su and Carreiras (2016) examined this assumption by contrasting grammatical gender and stereotypical gender simultaneously in Spanish noun phrases. Instead of the syntactic LAN<sup>11</sup> plus P600 Event-Related Potential (ERP) effects predicted by some previous studies (Gunter, Friederici & Schriefers, 2000; Barber & Carreiras, 2005; Molinaro, Vespignani & Job, 2008; O’Rourke & Van Petten, 2011; Caffarra et al., 2015), they observed long-lasting N400(-like) effects for grammatical violations which provide an opposite gender value compared with the stereotypical gender of the target head noun (see Table 4). The results not only confirmed the claim that stereotypical gender has a special status compared to other types of semantic knowledge (Contreras et al., 2012), but also supported the “stereotype over syntax” view when the two sources of gender information contrasted.

Critically, the results supported the “immediacy hypothesis”. Molinaro and colleagues concluded that the activation of stereotypical gender is mostly based on the lexical/semantic representation of a word stem (e.g., *miner-*) irrespective of the gender-informative morphemes presented in the determiner (e.g., *el/la*) and/or in a role noun (i.e. suffix as *-o* or *-a*). This may imply that in grammatical gender languages<sup>12</sup>, while the word stem of a stereotyped head noun leads the processing of a noun phrase, the morpho-syntactic gender information present in the determiner and role noun itself plays a limited role. If this is the case, the “immediacy hypothesis” may be generalized to genderless languages where morpho-syntactic gender morphemes are totally absent and similar N400 (-like) effects may also be expected. However, is the presence/absence of morpho-syntactic gender the only factor counting when evaluating a genderless language with no morpho-syntactic gender morphemes?

To be more specific, in a genderless language, such as Mandarin Chinese, the default stereotypical gender of a role noun is perceived via covert category of a generic role noun instead of relying on any linguistic gender-related cues (e.g., 矿工 /kuàng gōng/*miner*, male-biased). That is, each generic role noun has a default gender and it is commonly recognized by people who use this language (Farris, 1988).

<sup>11</sup> The LAN effect is a negative-going wave mostly observed in electrodes in the left anterior scalp peaking around 350 ms post stimulus onset. According to the literature, it is related to automatic parsing involving morpho-syntactic and rule-based decomposition processes (Hahne and Friederici, 2002; Barber and Carreiras, 2005; Hagoort, 2009; Molinaro et al., 2008a, 2008b, 2011).

<sup>12</sup> The linguistic coding of gender information varies by languages, and based on their grammatical features, languages are categorized into three groups: grammatical gender languages (morpho-syntactic gender agreement is required), natural gender languages (gender agreement is required for third person pronouns) and genderless languages (no morpho-syntactic gender features at all) (Stahlberg, Braun, Irmen & Sczesny, 2007).

Besides, Chinese gender markers are independent words (e.g., 男/nán/*male* or 女/nǚ/*female*) required to appear before (i.e., to the left of) a generic role noun in the case that the gender value is opposite to the default (e.g., ‘女’礦工/nǚ kuàng gōng/*female’ miner*). No change is made on the generic role noun itself and no gender agreement is required (which means that any combination of a gender marker and a generic role noun is grammatically correct). This is very different from Spanish (see previous chapter).

Under the fact that the gender markers and the way of perceiving stereotypical gender of a role noun are so different in Spanish and Mandarin Chinese, the goal of the present study is to investigate the immediacy hypothesis of stereotypical knowledge in Mandarin Chinese to see how the processing of stereotypical knowledge differs by languages of distinct gender marking systems.

**Table 4. The design reported in Molinaro, Su and Carreiras (2016): the lower part in each condition is the result when the condition is compared with the control condition.**

	<b>Stereotypicality congruent</b>	<b>Stereotypicality incongruent</b>
Grammaticality agreed	Los mineros (the <sub>DET/PL/MASC</sub> male miners)	Las mineras (the <sub>DET/PL/FEM</sub> female miners)
	Control condition	N400-like effect found
Grammaticality violated	Las mineros (the <sub>DET/PL/FEM</sub> male miners)	Los mineras (the <sub>DET/PL/MASC</sub> female miners)
	N400-like effect found	N400-like effect found

Gender in Chinese role nouns is expressed by covert category and/or definite gender markings<sup>13</sup> (Farris, 1988; Moser, 1997). Covert category depends on the conventional gender that is accepted through common practice in the society. The covert category is expressed via generic (unmarked) forms of role nouns, similar to the default stereotypical gender discussed in Carreiras and colleagues (1996). For instance, the generic 搬運工人/bān yùn gōng rén/*porter* typically belongs to the male class, even if a porter could be a male or a female in reality<sup>14</sup>. A gender modifier 女/nǚ/*female* or *woman* is added to mark gender overtly, as ‘女’搬運工人/nǚ bān yùn gōng rén/*female’ porter*. Similarly, the generic female biased 美甲師/měi jiǎ shī/*manicurist* belongs to the female class. The gender specific modifier 男/nán/*male* or *man* is added to make it male manicurist as ‘男’美甲師/nán měi jiǎ shī/*male’ manicurist*. Some authors have clearly stated that this gender configuration involves semantic features only (Farris, 1988). In a Chinese noun phrase, the modifiers (e.g., in an associative phrase, a classifier/measure phrase, a relative clause or adjectives) should appear before a head noun (i.e., they are pre-nominal, Li & Thompson, 1981). The structure of a Chinese noun phrase is shown in (11). Here, it can be appreciated that adjectives in Chinese noun phrases are to the left of the head noun, while, on the other hand, adjectives in romance languages tend to be post-nominal (Rudder, 2012), e.g., 幾 NUM 個 QUAN 女的 ADJ 搬運工人 NOUN in Chinese and *las*<sub>DET/PL/FEM</sub> *porter*<sub>NOUN</sub> *femeninas*<sub>ADJ</sub> in Spanish (the female porters in English).

<sup>13</sup> In Chinese, to explicitly express the gender of a role noun that is not of the covert category, a gender modifier (instead of a determiner) is required (Farris, 1998).

<sup>14</sup> If the gender of a person belongs to the covert category of a role noun, usually, there is no marking before the role noun. However, for the purpose of clarifying, the “male/female” can be added before a male/female-biased role noun.

(11) Determiner + Number + Classifier/Quantifier + Adjective + HEAD NOUN

那	幾	個	女的	搬運工人
nà	jǐ	gè	nǚ(-de)	bānyùngōngrén
that	several	Classifier	female(-Nominalizer)	porter

“those female porters”

In the present study, we investigate the activation of stereotypical gender during the processing of Chinese role nouns by manipulating the modifiers in the noun phrases. We used a 2 x 3 design, with two factors manipulated: Noun Stereotype (male-biased vs. female-biased) and Stereotypicality Congruity (congruent, incongruent and default) between the gender of a modifier and the covert category of a generic role noun within the same noun phrase. Two levels for Noun Stereotype were manipulated: male-biased (e.g., 搬運工人/*porter*) and female-biased (e.g., 美甲師/*manicurist*). We kept male and female stereotypes separate, since a number of studies have suggested a differential processing of these (Cacciari & Padovani, 2007; Siyanova-Chanturia, Pesciarelli & Cacciari, 2012; Irmen, 2007; Irmen, Holt & Weisbrod, 2010). Stereotypicality Congruity has three levels. Apart from the two gender specific modifiers, 男/*nán/male* and 女/*nǚ/female* that could be congruent or not with the stereotypical gender of the target noun, a third condition presenting the default stereotypical gender was included for the purpose of estimating the processing influence of the covert category without any interference of gender modifiers (i.e., no gender modifiers, only quantifiers, such as 幾個搬運工人/*jǐ gè bān yùn gōng rén/several porters* or 一些美甲師/*yī xiē měi jiǎ shī/some manicurists*). The three levels of Stereotypicality Congruity of a male-biased role noun are exemplified in (12) (see Table 5 for examples in sentences). The 的/*de* is a nominalization particle which is used depending on the literary style in Chinese (Li & Thompson, 1981). Different from previous Spanish study (Molinaro, Su, & Carreiras, 2016), no grammaticality manipulation was employed.

- (12) Stereotypical gender congruent: 幾個“男的”搬運工人  
(*several “male” porters*)  
Stereotypical gender incongruent: 幾個“女的”搬運工人  
(*several “female” porters*)  
Default stereotypical gender: “幾個”搬運工人  
(*“several” porters*)

**Table 5. Experimental conditions and example sentences presented in the experiment**

Condition		Example Sentences
Male-biased noun stereotypicality	Stereotype congruent	大樓門口，一些 <b>男的搬運工人</b> 正在將貨物搬上車。(40 sentences) (In front of the building, some <b>male porters</b> are loading the goods on the truck.)
	Stereotype incongruent	公司樓下，一些 <b>女的搬運工人</b> 正忙著將貨物搬下貨車。(40 sentences) Downstairs of the company, some <b>female porters</b> are busy for unloading the goods from the truck.
	Default stereotypical gender	工廠門口，一些 <b>搬運工人</b> 正忙著把貨物搬進倉庫。(40 sentences) In front of the factory, some porters are busy for moving the goods into the warehouse.
Female-biased noun stereotypicality	Stereotype congruent	在店裡，幾個 <b>女的美甲師</b> 正在討論最新流行的指甲彩繪。(40 sentences) In the store, several <b>female manicurists</b> are talking about the latest nail patterns.
	Stereotype incongruent	百貨公司裡，幾個 <b>男的美甲師</b> 正在推銷指甲油。(40 sentences) In the department store, several <b>male manicurists</b> are promoting nail polish.
	Default stereotypical gender	在休息室裡，幾個 <b>美甲師</b> 聚在一起喝咖啡聊是非。(40 sentences) In the resting lounge, several <b>manicurists</b> are drinking coffee and talking about gossips.

An ERP mismatch paradigm was employed. The advantage of the ERP approach resides in its dynamic temporal resolution time-locked to the target word of interest. Three components have previously been observed to correlate with the processing of the role noun: P200, N400 and P600 (see the last paragraph of section 2.1 ).

Here we will evaluate how the activation of stereotypical knowledge is affected by the linguistic properties of Chinese (a genderless language), and discuss the present findings in relation to our previous study in Spanish (a ‘morphological gender’ language). Based on Molinaro and colleagues (2016) and previous studies related to the immediate activation of stereotypical knowledge (see previous paragraph), when the gender marking of the target stimulus is inconsistent with the default stereotypical gender, this can trigger a N400 (-like) semantic effect (Kutas & Federmeier, 2011), either due to the stereotypical or to the grammatical conflict. Thus, a N400 (-like) semantic mismatch effect is expected for the incongruent condition as opposed to the default. Nonetheless, it is possible that gender marking before the target noun washes out the ‘immediate’ stereotype-congruity effect, with the consequence that no ERP differences may emerge.

Importantly, the main focus of the present study is on how gender markers influence the processing of stereotypical knowledge in a genderless language. Consequently, we will also focus on possible effects due to the presence or absence of gender markers. Given the lack of gender markers (a relevant cue for semantic gender recognition) before the generic role noun, we would expect that the default condition

should elicit differential brain activation at different processing stages: at the early perceptual phase (P200 effect, Liu, Perfetti & Hart, 2003; Lee, et al, 2007; Hsu, Tsai, Lee & Tzeng, 2009; Yum, et al, 2014; Hung & Schumacher, 2012; 2014; Su et al., 2016) until the late integration time intervals (P600 effect, Kaan, Harris, Gibson & Holcomb, 2000; Kaan & Swaab, 2003; van Herten, Kolk & Chwilla, 2005; Kuperberg, Caplan, Sitnikova, Eddy & Holcomb, 2006; Kuperberg, 2007; Van de Meerendonk, Kolk, Chwilla & Vissers, 2009; Brouwer, Fitz & Hoeks, 2012; Su et al., 2016; Burkhardt 2006; 2007; Hung & Schumacher, 2012; 2014; Wang & Schumacher, 2013; Burmester, Spalek & Wartenburger, 2014).

## 4.2 Methods

### 4.2.1 Participants

Forty-four native Chinese speakers (22 males, mean age: 23.89, range 20-33) were recruited from National Central University, Taiwan, and were paid for their participation. They were healthy, right-handed, with normal or corrected-to-normal vision. None of them reported prior history of neurological disorder. The design and execution of the experiment conformed to the ethical regulations of the Institute of Cognitive Neuroscience at National Central University in Taiwan. Informed consent was obtained from each participant.

### 4.2.2 Materials

The stereotypical role nouns tested in the present study were taken from a questionnaire containing 348 role nouns. All the role nouns on this questionnaire were generic without previous gender modifiers (e.g. 搬運工人/*bān yùn gōng rén/ porter*; 學生/*xué shēng/student*; or 美甲師/*měi jiǎ shī/manicurist*). Participants were required to evaluate the percentage of the composition of males and females to a given role noun and mark it on an 11-point Likert scale from 0% to 100% with the percentage of ten as the scale interval. The point of 0%-males/100%-females was set on the left and 100%-males/0%-females was set on the right. Fifty-seven (14 males, mean age: 21.58) college students rated this questionnaire in Tsing Hua University and Sun Yat-Sen University in Taiwan.

Forty male-biased and forty female-biased role nouns were selected as testing role nouns for the two levels of Noun Stereotype (male-biased vs. female-biased). Stereotypicality Congruity of the noun phrase (modifier + role noun) was manipulated in three levels as congruent, incongruent and the default. Both 男的/*nán de/male*, 女的/*nǚ de/female*, and quantified modifiers (e.g., 幾個/*jǐ gè/several*) were individually added before each of the 80 role nouns to form 240 noun phrases. 240 sentences containing the 240 noun phrases were created as experimental sentences. The structure of each sentence was similar. A segment of contextual information appeared first, and a statement containing the noun phrase as the subject of the sentence followed. The target word was always the role noun that was set on the third, fourth, or fifth position and never appeared in the final position of each sentence (see Table 2).

Comprehension yes/no questions appeared after thirty percent of the sentences (i.e., 72 comprehension questions) to ensure that participants did comprehend the sentences they read. The questions were only related to the content of the sentence and never related to any gender information of our interests.

### 4.2.3 Procedure

All the stimuli were presented in white letters on a black background. Each trial began with a fixation point “+” at the center of the screen for 500 ms, followed by a blank screen for 400 ms. Each sentence was presented word by word. The critical gender modifiers and the target role nouns were presented separately. Each word was presented for 400 ms followed by a blank screen for 400 ms. According to some previous studies (Jiang & Zhou, 2009; Ye et al., 2007; Su et al., 2016), the 400 ms (word) + 400 ms (blank) stimuli presentation is natural and comfortable for Chinese readers. A variable inter-trial time interval (from 1700 ms to 3000 ms) appeared after each sentence.

Participants sat in a sound-attenuated cabin and were instructed to read each sentence silently and carefully. The task was to answer yes/no comprehension questions by pressing one of the pre-designated buttons. The response buttons were always the same across participants.

A practice section with 12 trials was conducted before the main experiment. The main experiment was arranged in six blocks with five breaks. Each block contained 40 sentences. All the 240 sentences were presented randomly employing Presentation experimental software (<http://www.neurobs.com>).

### 4.2.4 Data Recording and Analysis

Continuous EEG data were recorded by SynAmp2 (NeuroScan) with a 32 electrodes cap and the electrodes were positioned according to the 10-20 system. The impedance was kept below 5 $\Omega$  in each electrode. The sampling rate (A/D) was 1000 Hz. The on-line reference electrode was the left mastoid (M1) and the right mastoid (M2) was also recorded. The signals were amplified with a bandpass of 0.05 - 200 Hz. The ground electrode was set between FPZ and FZ. HEOGs were placed at the outer canthi of the eyes and VEOGs were placed above and below the left eye in a bipolar montage.

The EEG raw data were re-referenced to the average activity of M1 and M2. Signal was bandpass filtered between 0.1 Hz and 30 Hz. Epochs of interest lasted from -200 ms before the onset of the target role nouns to 1000 ms post-stimulus onset. Baseline correction was computed considering the 200 ms activity prior to the onset of the target role nouns. Trials with artifacts, such as eye blinks or saccades, or with activity exceeding  $\pm 120 \mu\text{V}$  were rejected. As a result, 3.21% of the trials were removed due to artifact rejection.

Data analysis focused on the mean voltage of each electrode within a certain time interval of interest across all the participants. Based on the findings in previous ERPs studies, P200, N400 and P600 components were used to index the processing differences. The time intervals involved were chosen based on the literature and visual inspection on the ERPs waveforms.

Repeated measures ANOVAs were separately employed for groups of electrodes in the midline region (anterior: Fz; central: Cz; and posterior: Pz) and the lateral scalp regions (left hemisphere: anterior, Fp1, F3, F7, FC5; central, FC1, C3, CP1, T7; posterior, CP5, P3, P7, O1; right hemisphere: anterior, Fp2, F4, F8, FC6; central, FC2, C4, CP2, T8; and posterior, CP6, P4, P8, O2). A three-way repeated measures ANOVA was employed for electrodes in the midline region considering Noun Stereotype (male-biased vs. female-biased), Stereotypicality Congruity between the modifier and the role noun (congruent, incongruent and default), and the topographical factor, Latitude (anterior, central and posterior). For electrodes in the

lateral scalp regions, a four-way repeated measures ANOVA was used including the two experimental factors and the two topographical factors: Latitude (anterior, central and posterior) and Hemisphere (left vs. right). Effects related to the topographical factors were reported only when they interact with the experimental factors. If there was any effect involving Stereotypicality Congruity, *post hoc* pairwise comparisons between each two levels of this factor were carried out to look for differences between each two of the three manipulations. The Greenhouse-Geisser corrected *p*-value was used if the degrees of freedom were larger than one or for violations of sphericity.

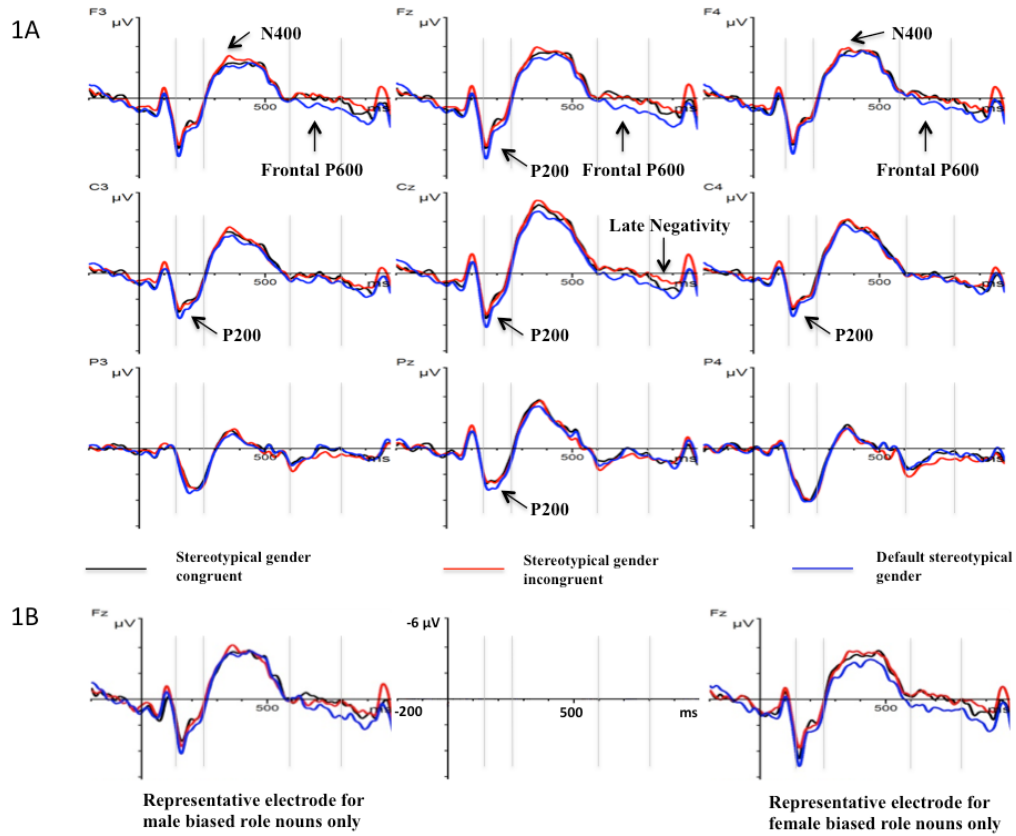
## **4.3 Results**

### **4.3.1 Comprehension Questions**

The mean accuracy for comprehension questions was .93 ranging from .89 to .98, showing that participants did comprehend the sentences they read.

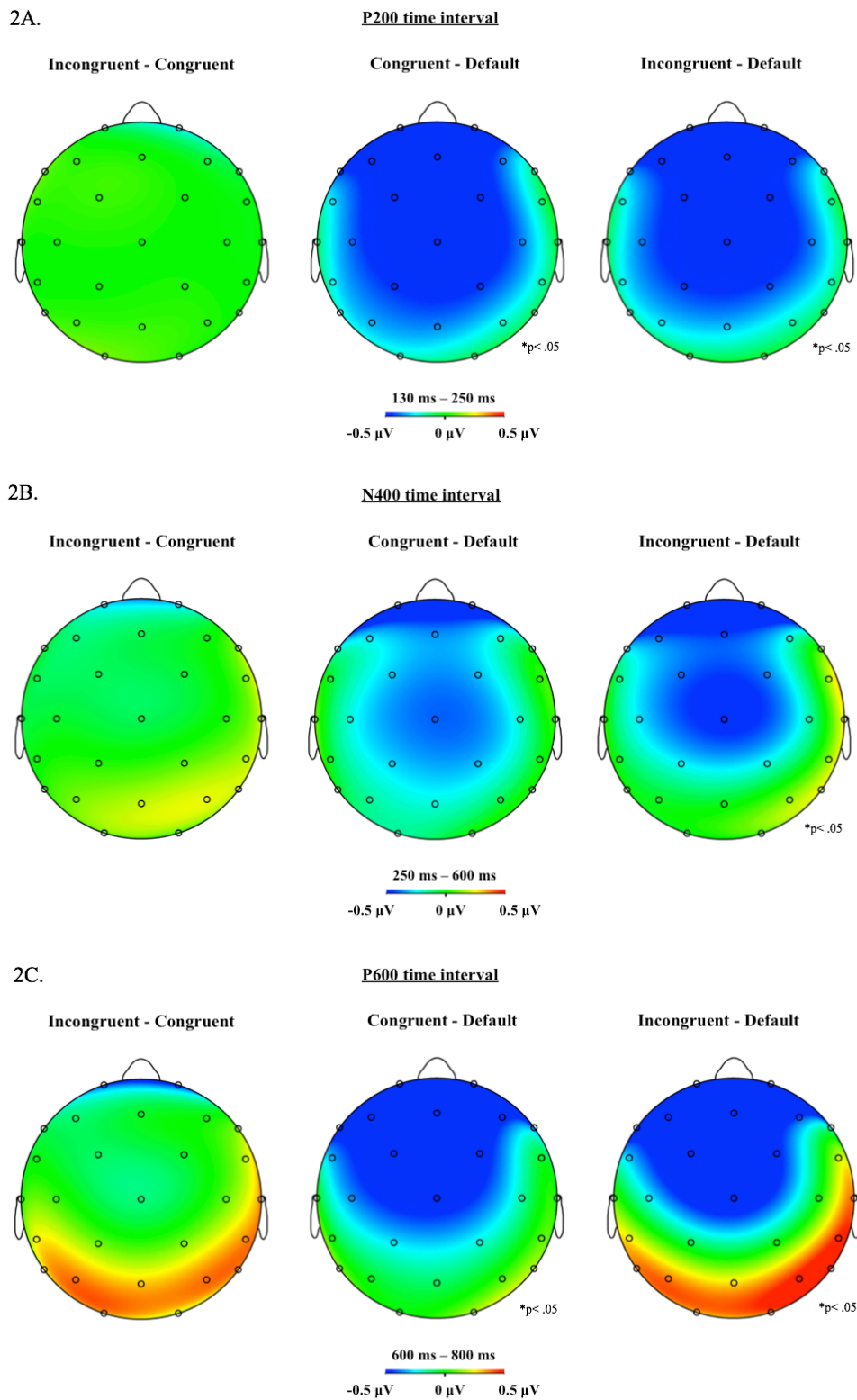
### **4.3.2 ERPs on the Role nouns**

Figure 6, 1A represents grand average of the ERP waveforms elicited by the three manipulated conditions of Stereotypicality Congruity (without taking into account Noun Stereotype) in the representative electrodes. The visual inspection on the overall waveforms suggested that the time intervals of interests were 130–250 ms for the P200, 250–600 ms for the N400, and 600–800 plus 800–1000 ms for the P600 component respectively.



**Figure 6. 1A: ERPs time-locked to the target words of the three experimental conditions of Stereotypical Gender Congruity at nine representative electrodes in the frontal, central and posterior scalp regions. 1B: Representative electrode (Fz) in the mid-anterior region respectively for male biased role nouns and female biased role nouns.**





Created by Paint X

Figure 7. Topographic distribution of the pair-wised congruity effect (incongruent minus congruent or congruent/incongruent minus default) respectively in the P200 (2A), N400 (2B) and P600 (2C) time intervals.

#### 4.3.2.1 The P200 Time Interval (130–250 ms)

The repeated measure ANOVA on the brain activity showed significant main effects of Stereotypicality Congruity in the midline [ $F(2, 86) = 4.80$ ,  $MSE = 5.05$ ,  $p = .011$ ] and lateral scalp [ $F(2, 86) = 4.30$ ,  $MSE = 4.66$ ;  $p = .017$ ] regions. *Post hoc* comparisons between each two levels of Stereotypicality Congruity revealed significant differences only when the default was respectively compared with the congruent [midline:  $F(1, 43) = 6.19$ ,  $MSE = 3.17$ ,  $p = .017$ ; lateral scalp:  $F(1, 43) = 5.05$ ,  $MSE = 3.03$ ,  $p = .030$ ] and the incongruent [midline:  $F(1, 43) = 7.56$ ,  $MSE = 2.20$ ,  $p = .009$ ; lateral scalp:  $F(1, 43) = 7.06$ ,  $MSE = 2.09$ ,  $p = .011$ ] conditions. No such effects were found for the comparisons between the congruent and incongruent manipulations [midline:  $F(1, 43) = .05$ ,  $MSE = 2.21$ ,  $p = .817$ ; lateral scalp:  $F(1, 43) = .00$ ,  $MSE = 1.86$ ,  $p = .955$ ]. Thus, in this P200 time interval, the default was found to elicit more positive amplitude than other two congruity-related conditions (see Table 6A and Figure 7, 2A).

An interaction between Noun Stereotype and topographic factor [ $F(2, 86) = 5.18$ ,  $MSE = .57$ ,  $p = .019$ ] emerged in the midline region. *Post hoc* tests showed a significant difference between male-biased and female-biased role nouns in the mid-anterior region [ $F(1, 43) = 5.96$ ,  $MSE = .76$ ,  $p = .019$ ; male-biased:  $2.15 \mu V$ ; female-biased:  $2.61 \mu V$ ], showing that female-biased role nouns elicited more positive brain activity than male-biased ones (See Figure 6, 1B).

#### 4.3.2.2 The N400 Time Interval (250–600 ms)

The repeated measures ANOVA showed no relevant effects for electrodes in the midline region. The ANOVA for lateral scalp regions showed an interaction between Stereotypicality Congruity and Latitude [ $F(4, 172) = 3.01$ ,  $MSE = .86$ ,  $p = .037$ ]. *Post hoc* comparisons between each two levels of Stereotypicality Congruity showed a significant difference for the incongruent condition in the anterior region when compared with the default [ $F(1, 43) = 4.43$ ,  $MSE = 1.86$ ,  $p = .041$ ] (see Table 6B and Figure 7, 2B). A N400-like mismatch effect was observed in the anterior region for the incongruent condition (incongruent < default) and no differences emerged for other two comparisons (congruent vs. incongruent; default vs. congruent).

#### 4.3.2.3 The P600 Time Interval (600–800 ms and 800–1000 ms)

At the early P600 time interval, the repeated measures ANOVA showed interactions between Stereotypicality Congruity and Latitude in the midline [ $F(4, 172) = 6.78$ ,  $MSE = .89$ ,  $p < .001$ ] and lateral scalp [ $F(4, 172) = 10.93$ ,  $MSE = 1.68$ ,  $p < .001$ ] regions. *Post hoc* comparisons showed significant differences in the anterior region when the default was compared with the congruent conditions [midline:  $F(1, 43) = 7.83$ ,  $MSE = 1.76$ ,  $p = .008$ ; lateral scalp:  $F(1, 43) = 7.05$ ,  $MSE = 3.98$ ,  $p = .011$ ] and incongruent [midline:  $F(1, 43) = 4.07$ ,  $MSE = 2.50$ ,  $p = .050$ ; lateral scalp:  $F(1, 43) = 10.25$ ,  $MSE = 3.32$ ,  $p = .003$ ], but no such effects were found between the congruent and incongruent conditions [midline:  $F(1, 43) = .14$ ,  $MSE = 1.98$ ,  $p = .710$ ; lateral scalp:  $F(1, 43) = .09$ ,  $MSE = 3.32$ ,  $p = .767$ ] (see Table 6C and Figure 7, 2C).

At the late P600 time interval, significant interactions between Stereotypicality Congruity and Latitude were revealed in the midline [ $F(4, 172) = 7.21$ ,  $MSE = 1.00$ ,  $p < .001$ ] and lateral scalp [ $F(4, 172) = 11.40$ ,  $MSE = 2.24$ ,  $p < .001$ ] regions. The planned *Post hoc* comparisons showed significant effects that emerged for the incongruent condition in the mid-central brain region when it was compared with the congruent condition [ $F(1, 43) = 4.84$ ,  $MSE = 1.97$ ,  $p = .033$ ] and the default [ $F(1, 43) =$

8.75, MSE= 2.11,  $p= .005$ ]. Significant differences were found for the default condition in the lateral anterior region as opposed to the congruent [ $F(1, 43)= 4.63$ , MSE= 5.42,  $p= .037$ ] and incongruent [ $F(1, 43)= 16.92$ , MSE= 4.13,  $p< .001$ ] conditions (see Table 6D).

In the early P600 time interval (600–800 ms), while the default elicited more positive brain activation in the anterior region than other two congruity-related conditions, no differences emerged between the two congruity-related conditions. At the later stage (800–1000 ms) right after the early P600 time interval, the incongruent condition showed more negative brain activation as opposed to the congruent and the default in the mid-central brain region.

**Table 6. Mean amplitudes and *F* values resulting from Post hoc pairwise comparisons between each two manipulated conditions of STEREOTYPICAL GENDER CONGRUITY in the specific topographic regions respectively in the P200, N400 and P600 time intervals.**

A.	P200 time interval (130 ~ 250 ms)					
	Mean amplitude ( $\mu\text{V}$ )			<i>F</i> values ( <i>df</i> = 1, 43)		
	Congruent	Incongruent	Default	Congruent vs. incongruent	Default vs. congruent	Default vs. incongruent
Midline	2.02	2.07	2.57	— <sup>a</sup>	6.19*	7.56**
Lateral scalp	1.72	1.73	2.06	— <sup>a</sup>	5.05*	7.06*
B.	N400 time interval (250 ~ 600 ms)					
	Mean amplitude ( $\mu\text{V}$ )			<i>F</i> values ( <i>df</i> = 1, 43)		
	Congruent	Incongruent	Default	Congruent vs. incongruent	Default vs. congruent	Default vs. incongruent
Lateral anterior	-1.68	-1.75	-1.32	— <sup>a</sup>	2.54	4.43*
Lateral central	-1.92	-1.90	-1.73	— <sup>a</sup>	— <sup>a</sup>	— <sup>a</sup>
Lateral posterior	0.24	0.35	0.30	— <sup>a</sup>	— <sup>a</sup>	— <sup>a</sup>
C.	P600 time interval (600 ~ 800 ms)					
	Mean amplitude ( $\mu\text{V}$ )			<i>F</i> values ( <i>df</i> = 1, 43)		
	Congruent	Incongruent	Default	Congruent vs. incongruent	Default vs. congruent	Default vs. incongruent
Midline anterior	-0.12	-0.01	0.67	— <sup>a</sup>	7.83**	4.07*
Lateral anterior	0.04	-0.05	0.83	— <sup>a</sup>	7.05*	10.25**
Lateral central	0.18	0.30	0.46	— <sup>a</sup>	1.66	— <sup>a</sup>
Lateral posterior	0.48	0.78	0.40	3.23	— <sup>a</sup>	3.75
D.	Late P600 time interval: 800 ~ 1000 ms					
	Mean amplitude ( $\mu\text{V}$ )			<i>F</i> values ( <i>df</i> = 1, 43)		
	Congruent	Incongruent	Default	Congruent vs. incongruent	Default vs. congruent	Default vs. incongruent
Midline anterior	0.83	0.35	1.47	2.41	4.00	9.16**
Midline central	0.78	0.12	1.04	4.84*	— <sup>a</sup>	8.75**
Midline posterior	0.73	0.56	0.65	— <sup>a</sup>	— <sup>a</sup>	— <sup>a</sup>
Lateral anterior	0.76	0.26	1.52	3.15	4.63*	16.92**
Lateral central	0.81	0.50	0.94	2.43	— <sup>a</sup>	2.89
Lateral posterior	-0.13	0.06	-0.18	— <sup>a</sup>	— <sup>a</sup>	1.52

—<sup>a</sup> *F* < 1; \* *p* < .05; \*\* *p* < .01

## 4.4 Discussion

The present study was aimed at testing online activation of stereotypical knowledge in a genderless language, i.e., Mandarin Chinese here. The results showed an attention-related P200 and an integration-related P600 effect for the default as opposed to the other two gender-specified conditions, and a N400-like effect revealed for the incongruent condition when compared with the default (see Figure 7, 2B). Also, a very late effect (likely 'post-semantic' processing, 800–1000 ms) emerged for stereotypically incongruent noun phrases when compared with the default and the congruent ones. This is similar to the long-lasting negativity reported in Molinaro and colleagues (2016) for stereotypically incongruent nouns. Based on the results, it seems that the immediacy of stereotypical knowledge is similar across languages, but the use of covert category, along with the presence/absence of gender markers modulates the immediate effect of stereotypical knowledge in Chinese.

### 4.4.1 Processing Modulation Given by Gender Marking: The P200

As discussed in the introduction, Chinese speakers make use of covert category to define the gender of generic role nouns. In general, the covert category of a generic role noun is conveyed by an unmarked form, i.e., without any gender specific information ahead of it. According to Farris (1988), “*the unmarked category is the culturally supposed "usual" case*” (see Greenberg, 1966), and “*has "zero" expression*” (Roman Jakobson, 1939) (cf., Farris, 1988, pp. 282-283). If the gender of a referent conflicts with its covert category, a gender specific modifier should be added before this role noun. We focus our discussion on two aspects: the gender markedness in Chinese noun phrases (i.e., marked vs. unmarked) and the information status (i.e., new vs. given, Burkhardt, 2006; 2007) of gender while participants process role nouns.

When the two gender-marked conditions were compared, the result- a relatively late effect between the congruent and the incongruent conditions (as compared to our previous study in Spanish) is very likely due to the presence of pre-nominal gender modifiers defining the gender for the forthcoming role noun. However, when the comparison was between the marked and unmarked forms, an increased N400-like effect emerged only for the incongruent condition (marked) and an increased P200 and a P600 effect emerged for the default (unmarked)<sup>15</sup>. Without gender modifiers, it seems that participants need more effort to process the generic role noun at the (P200) word recognition phase (Liu, Perfetti & Hart, 2003; Lee, et al, 2007; Hsu, Tsai, Lee & Tzeng, 2009; Yum, et al., 2014; Hung & Schumacher, 2012; 2014; Su et al., 2016) and at the later (P600) integration phase.

In terms of the greater effort participants make to process the default, the “given-new” distinction for inferential processing of Determiner phrases (DPs, determiner + noun) suggested by Burkhardt (2006, 2007) may account for the effects we report here (see also the fourth paragraph in section 2.4.1). In general, the P200 effect reflects attention-related feature analysis (Luck and Hillyard, 1994) and has been associated to orthographic processing in language studies (Dambacher et al., 2006,

<sup>15</sup> It could be argued that the later effects observed after the P200 time interval (in the N400 and P600 intervals) reflect a carry-over of the more positive deflection for the unmarked condition compared to the marked ones. Although we acknowledge that this could reflect a basic ambiguity of the ERP methodology in which later effects are baseline-affected by earlier effects, we also observe that the unmarked condition triggered stronger effects in the P200 and the P600 time intervals (see Figure 2). Consequently, it is worth considering the P600 as a late additional processing cost that did not dissipate after the initial visual-attention P200 effect.

Barber et al., 2004; Carreiras et al., 2005, Liu et al., 2003; Lee et al., 2007; Hsu et al., 2009; Yum et al., 2014, Hung and Schumacher, 2012, 2014). In Hung and Schumacher's studies (2012; 2014), P200 effects were found to be more positive for "new" as compared with "given" information (new > given, 2014; topic-shift > topic-continuity, 2012). The authors claimed that the P200 effect reflects early perceptual costs of processing new information, depending on the availability from context. The same could be true here for the availability of gender cues from context. In the present study, the appearance of a role noun is expected after a gender modifier in each sentence. On the other hand, with no previous gender modifier the unmarked role nouns convey new discourse-relevant gender information as opposed to the marked ones. The brain would need more effort to perceptually process the unmarked role nouns to retrieve the default stereotypical knowledge and to integrate (or wrap-up) all the related information. The more positive P200 effect for the default condition very likely reflects such efforts in attention-related perceptual processing. The topographic distribution of this P200 effect is located in the anterior region, similar to previous studies related to Chinese discourse inference (Hung & Schumacher, 2012; 2014; Su et al., 2016). In fact, in a Chinese reflexive pronoun resolution study, an attention-related P200 effect was found for the gender-specific reflexive pronoun, 她自己 /*herself*, as compared with the default, 他自己 /*himself*, (Su et al., 2016). And interestingly, the gender difference between *herself* and *himself* is similar to that between the marked (specific) and unmarked (default) role nouns in the present study, although one is encoded in the radical of a character (see Su et al., 2016) and the other is determined by the presence/absence of gender modifiers.

#### 4.4.2 Processing Modulation Given by Gender Marking: The P600

The P600 effect observed for the default condition may reflect the increased cost of updating information in discourse memory (Burkhardt 2006; 2007; Hung & Schumacher, 2012; 2014; Wang & Schumacher, 2013; Burmester, Spalek & Wartenburger, 2014) among other possible explanations. First, no grammatical violations were involved in each of the sentences presented in the default condition and only semantic features (i.e., gender modifiers, Farris, 1988) were entailed in the comparisons. Second, the sentences presented in the default condition did not involve any semantic-related anomalies. The sentence structure and thematic role were always the same across the three conditions. The only difference for the default condition was the absence of gender modifiers. Thus, any explanation referring to some form of semantic anomaly (Van de Meerendonk, Kolk, Chwilla & Vissers, 2009; Van de Meerendonk, Kolk, Vissers & Chwilla, 2010; Brouwer, Fitz & Hoeks, 2012) would not be applicable to the present results. Consequently, the effect cannot possibly reflect syntactic/conceptual-related integration difficulties (Kaan, Harris, Gibson & Holcomb, 2000; Kaan & Swaab, 2003; Carreiras, Salillas & Barber, 2004; van Herten, Kolk & Chwilla, 2005; Kuperberg, Caplan, Sitnikova, Eddy & Holcomb, 2006; Kuperberg, 2007).

On the other hand, robust P600 effects for new inferential noun phrases (NPs) have been reported in studies related to discourse processing (Hung & Schumacher, 2012; 2014; Wang & Schumacher, 2013; Burmester, Spalek & Wartenburger, 2014), with the P600 effects interpreted as reflecting the increased demands of updating the information in discourse memory. Accordingly, in the present study, target role nouns with/without gender modifiers may represent given/new inferential NPs. While participants' sensitivity towards gender is reduced by the presence of previous gender

modifiers the absence of gender modifiers in the default condition would result in some extra cost during information processing. Thus, in addition to the increased cost required for word recognition at the early perceptual phase, the processing of Chinese default role nouns also requires more cognitive resources for information processing of discourse-related information in memory at the later integration phase, as opposed to gender-specific information. Note that this P600 effect is anteriorly distributed (frontal P600), in line with the information-updating hypotheses in some previous studies (Friederici, Hahne & Saddy, 2002; Kaan & Swaab, 2003). We would like to note that this finding may contrast with the general idea that processing of unmarked information requires less cognitive resources (see Deutsch & Bentin, 2001 for similar P600 effects for the marked condition). The distinct gender marking system and the word order of Chinese noun phrases (i.e., gender modifier + role noun) are the keys to this difference. In contrast to languages employing gender-related bound morphemes as suffixes of target words, Chinese gender modifiers are individual characters presented before role nouns. Here, processing of stereotypical knowledge possibly diverges because of distinct gender marking systems across languages.

#### 4.4.3 Stereotypical congruence: The case of Chinese

Despite the linguistic markedness and new-given information status that can account well for the P200 and P600 effects in the default condition, these cannot explain the N400-like effect found for the incongruent condition. As shown in Table 3B, this N400-like effect reflects pure semantic mismatch if compared with default stereotypical knowledge. This result may support the claim made by Deutsch and Bentin (2001) that the N400 effect is not influenced by linguistic markedness but by semantic anomaly status of the stimulus. Therefore, a pure semantic anomaly interpretation (Kutas and Federmeier, 2011, for a review) can account for the N400-like effect observed here, consistent with previous studies related to the processing of stereotypical knowledge (Van Berkum, van den Brink, Tesink, Kos & Hagoort, 2008; White, Crites, Taylor & Corral, 2009; Irmén, Holt & Weisbrod, 2010; Wang, Yang & Zhao, 2010; Siyanova-Chanturia, Pesciarelli & Cacciari, 2012; Spanish: Molinaro, Su & Carreiras, 2016). However, different from the typical N400 effect with central-to-posterior topographic distribution, the anteriorly distributed N400 effect for the incongruent condition in the present study is consistent with the topographic distribution of the N400 effect in stereotype processing as reported in Molinaro, Su & Carreiras (2016, see also White et al., 2009). Thus, the current study confirms that the processing of social knowledge is qualitatively different from that of general semantic concepts (Contreras, Banaji & Mitchell, 2012).

On the other hand, the lack of differences in early measures (i.e., until the N400 interval) found for the marked conditions (congruent vs. incongruent) are similar to the findings reported in Kreiner, Sturt & Garrod (2008, Experiment 2) in which cataphoric pronoun resolution structure was employed. In this latter experiment, the gender of a reflexive pronoun is expressed at the beginning of an utterance and no effect was recorded on the target stereotypical role nouns. Please see example (13).

(13) After reminding *himself* / \**herself* about the letter, the minister immediately went to the meeting at the office.

(example taken from Kreiner, Sturt & Garrod, 2008, p.244)

According to these authors, the activation of stereotypical gender depends on the context, and only when the gender of a certain stereotypical role noun is not clear, does the automaticity/immediacy of stereotypical knowledge trigger an observable effect. Thus, different from definitional gender that is lexically encoded, stereotypical gender is context-dependent and inferential at the discourse level (for detailed discussion, please see Kreiner, Sturt & Garrod, 2008). Similarly, in the present study (although the linguistic context is a noun phrase rather than a cataphoric structure), the results show no early (immediate) effect between the congruent and incongruent conditions when the gender modifier appears before a generic role noun. Given that the early activation of stereotypical knowledge can be silenced by the presence of a gender modifier, we may infer that Chinese stereotypical gender is not encoded at the lexical level and its processing can be more sensitive to context compared to other languages. In contrast to Spanish, where stereotypical knowledge is lexically represented in the word stem (e.g., *miner-* /*miner* in English, male stereotyped, Molinaro et al., 2016), stereotypical knowledge in Chinese is inferential and context-dependent (and this can be true for languages with no morpho-syntactic gender on role nouns, as proposed for English by Kreiner, Sturt and Garrod, 2008). Such contextual information available for gender-specification may be social, discourse, or textual.

The inferential nature of stereotypical knowledge in Chinese is confirmed by the late mismatch effect in a very late time interval (as evidence by the ‘post-semantics’ mismatch effect found for the incongruent as opposed to the other two conditions). This late negativity for incongruent items could possibly be linked to the long-lasting negativity reported for stereotypes mismatched in Molinaro et al. (2016). Importantly, while the effect started relatively early in Spanish (~300 ms), here it shows a much later onset.

## **4.5 Conclusion**

The present study investigated how the immediacy of stereotypical knowledge is manifested in a genderless language (e.g. Mandarin Chinese here) with totally different gender marking system as contrast to Spanish (a grammatical gender language. Consistent with previous studies related to the activation of stereotypical knowledge, an N400 effect was found when the gender of a referent does not match the default stereotypical gender of a noun phrase (White, Crites, Taylor & Corral, 2009; Siyanova-Chanturia, Pesciarelli & Cacciari, 2012; Molinaro, Su & Carreiras, 2016). A late (post-semantic) negativity was triggered by stereotypicality-incongruent noun phrases (as opposed to congruent and default), resembling the long lasting negativity reported in Molinaro, Su and Carreiras (2016) but with a later onset. Given the absence of a clear gender modifier beforehand, processing Chinese default stereotypical gender (i.e., the covert category of a generic role noun) requires more effort at early word recognition (P200) and late information updating (P600) phases as opposed to processing a gender specified role noun. Thus, different from previous studies on grammatical gender languages, the present findings show that stereotype processing is sensitive to the linguistic cues and structure available in a target language. Stereotypical knowledge is inferential and can be modulated by the presence of a gender modifier (before a role noun) in Chinese. Probably, the activation mechanism of stereotypical gender is language-specific as reported here (in Chinese) or in Spanish (Molinaro et al., 2016). Future cross-linguistic studies or studies on bilingual processing may shed additional light on this issue.



## **CHAPTER 5**

### **Overall Discussion of the Present Dissertation**

## 5.1 Summary of the Results in Each Experiment

The present dissertation is aimed at investigating the processing of gender information in languages of distinct gender marking systems, mainly focused on anaphoric pronoun resolution in a genderless language (Mandarin Chinese, Experiment I) and the processing of stereotypical knowledge in a grammatical gender language (Spanish, Experiment II) and a genderless language (Mandarin Chinese, Experiment III).

In Experiment I (Chapter Two), the type of antecedents and the asymmetry of gender specificity encoded in Chinese third person reflexive pronouns were investigated with the structure requiring reflexive pronoun resolution. The results showed: first, no significant differences revealed between the two types of antecedents. Second, an attention-related P200 effect (她自己/*herself* > 他自己/*himself*) and a semantic-related N400 effect (她自己/*herself* < 他自己/*himself*) were found for 她自己/*herself*, supporting the assumption that 他自己/*himself* is a default pronoun. Third, a P200 congruity effect for 他自己/*himself* (congruent > incongruent) and a P600 congruity effect for 她自己/*herself* (incongruent > congruent) emerged. This dissociation provided consistent evidence on the asymmetry of gender specificity in Chinese third person pronouns and suggested the distinct time courses of gender processing during reflexive pronoun resolution.

In Experiment II (Chapter Three), how stereotypical knowledge interacts with grammatical gender is evaluated online during the comprehension of Spanish stereotypical noun phrases (determiner + role noun). When stereotypical gender and grammatical gender are simultaneously contrasted in the noun phrases, N400 effects were found in all the three gender conflicts (stereotypical: las<sub>FEM</sub> mineras<sub>FEM</sub>/the<sub>FEM</sub> female miners; grammatical: las<sub>FEM</sub> mineros<sub>MASC</sub>/the<sub>FEM</sub> male miners; and double mismatch: los<sub>MASC</sub> mineras/the<sub>MASC</sub> female miners) when compared to the control (los<sub>MASC</sub> mineros<sub>MASC</sub>/the<sub>MASC</sub> male miners). The results showed that the default stereotypical knowledge leads the comprehension of a stereotyped noun phrase regardless of its morpho-syntactic gender morphemes, thus confirmed the immediacy hypothesis. This finding added on the evidence for the special status of stereotypical knowledge since social knowledge is claimed to be different from general semantic meanings (Contreras, Banaji & Mitchell, 2012).

In Experiment III (Chapter Four), the activation of stereotypical knowledge is observed online during the processing of Chinese stereotyped noun phrases (gender modifier + role noun). The results showed: first, a N400 semantic mismatch effect emerged for the incongruent condition only when it was compared with the default (i.e., the generic role noun without any preceding gender modifier). Second, an attention-related P200 and an integration-related P600 effects were found for the default condition. Third, a late (post-semantic) negativity was reported for stereotypicality incongruent noun phrases as opposed to the congruent and the default conditions. Though supporting the immediacy hypothesis (Carreiras, Garnham, Oakhill & Cain, 1996), the results pointed to a different processing mechanism and a distinct linguistic representation of stereotypical knowledge (i.e., inferential) in a genderless language with totally different gender marking system as contrast to a grammatical gender language.

## **5.2 *The Processing of Gender Information during Anaphoric Pronoun Resolution across Languages with Distinct Gender Marking Systems***

Along the completion of Experiment I, the entire picture of gender information processing (both definitional and stereotypical) in the three language groups (e.g., genderless languages, natural gender languages and grammatical gender languages) is complete. Some representative ERP studies (with one antecedent) belonging to each of the three language groups are listed in Table 7. The connection between the two-stage model and the relative ERP findings for gender processing during anaphor resolution are discussed in an incremental mode by each language group in the following section.

### **5.2.1 *The ERP Correlates Reported in Studies Involving Anaphor Resolution across Languages with Distinct Gender Marking Systems***

During anaphor resolution, only when there is an unambiguous mismatch on gender, the effects can be observed and the ERP components related to anaphor resolution are P200, N400 and P600. According to the studies shown in Table 7, some points can be integrated: first, the effect (P200) related to gender specificity of pronouns is reported only in Chinese, a genderless language. Second, anaphoric pronoun resolution for antecedents of both definitional and stereotypical role nouns are similar in genderless language and natural gender language, but not in grammatical gender language. Third, the P600 effect is robust for gender inconsistent anaphoras with antecedents of definitional role nouns across all the three language groups (except for the memory oriented P600 effect in Qiu and colleagues, 2012), and the nature of each individual P600 effect corresponds to the gender marking system of the testing language (the P600 effect is semantics-based in a genderless language and is syntactical in natural gender and grammatical gender languages). Fourth, the N400 effects are reported in genderless language (i.e., Mandarin Chinese, Experiment I, Su, Molinaro, Gillon-Dowens, Tsai, Wu & Carreiras, 2016) and grammatical gender language (i.e., Schmitt, Larmers & Münte, 2002; German, Irmen, Holt & Weisbrod, 2010) and they are respectively related to gender specificity of pronouns and antecedent's gender explicitness.

**Table 7. List of ERP Studies on the Processing of Gender Information during Anaphoric Pronoun Resolution**

Authors	Language	Language group	antecedent's gender type	Number of antecedent	main focus of the study	ERP results on pronouns with inconsistent gender	Anaphors used in the experiment	Interpretations of the results given by the authors
Qiu, Swaab, Chen & Wang, 2012	Chinese	Genderless	definitional	1	how working memory related to gender processing	N400 for inconsistent gender in short distance and P600 inconsistent lexical gender for long distance	singular third person pronouns used (he and she)	the processing of gender information decays as long as the distance between an antecedent and its anaphor gets longer, but the processing is "semantic-based"
Xu, Jiang & Zhou, 2013	Chinese	Genderless	definitional	1	pronoun resolution on gender and number	gender mismatch elicited earlier (singular) and larger (singular and plural) P600 as compare to number mismatch P600 (plural earlier and singular later); P600 for double mismatch is identical to gender mismatch	third person pronouns used (singular form in Exp 1 and plural form in Exp 2)	the P600 effects may represent either "semantic-based" processing or integration difficulty in general
Su, Molinaro, Gillon-Dowens, Tsai, Wu & Carreiras, 2016	Chinese	Genderless	definitional and stereotypical	1	how the gender of antecedent interact with the gender specificity of pronouns	P200 attention-related congruency effect found for himself; N400 for herself as compared to himself and P600 for inconsistent gender	singular reflexive third person pronouns used (general himself and specific herself)	the P600 effects reflect the semantic-based integration difficulty; there might exist a one-stage processing when the gender specificity of a pronoun can be applied to every individual
Osterhout, Bersick & McLaughlin, 1997	English	Natural gender	definitional and stereotypical	1	how the gender of antecedent interact with the lexical/syntactic gender of pronouns	P600 for lexical/syntactic gender violations	singular reflexive third person pronouns used (himself and herself)	gender information is encoded in the grammar and results in a syntactic processing
Kreiner, Mohr, Kessler & Garrod, 2009	English	Natural gender	definitional and stereotypical	1	how the linguistic context affect gender processing (anaphoric and cataphoric)	P600 for lexical/syntactic gender violations in anaphoric pronoun resolution	singular reflexive third person pronouns used (himself and herself)	regarding anaphoric pronoun resolution, the results are similar to what is found in Osterhout and colleagues (1997) as syntactic
Schmitt, Lamers & Münte, 2002	German	Grammatical gender	definitional	1	the interaction between semantic gender and syntactic gender	P600 for all syntactic gender violations; N400 for non-diminutive	singular third person pronouns used (he, she and it)	anaphoric pronoun resolution is basically syntactic, but can interact with semantic gender
Irmen, Holt & Weisbrod, 2010	German	Grammatical gender	stereotypical	1	how stereotypical gender interacted with the lexical gender of anaphoric noun phrases	N400 for stereotypically male antecedent as compared to stereotypically female antecedents; P600 for lexical gender incongruency between antecedent and anaphoric NP	plural anaphoric noun phrases are used (these men, these women, or the people)	stereotypical gender information is processed at the N400 time interval; the integration of stereotypical gender and lexical gender is processed later at the P600 time interval

### 5.2.2 The ERP Correlates and the Two-stage Model for Anaphoric Pronoun Resolution across Languages with Distinct Gender Marking Systems

The processing mechanism for anaphor resolution is often explained along the lines of the two-stage model in which the linking of related information is processed at the bonding stage and the integration of the collected information is operated at the resolution stage (Garrod and Sanford, 1994; Garrod & Terras, 2000, see Chapter 1, section 2.3). On the other hand, the correlations between the two-stage model and anaphor resolution studies employing ERP mismatch paradigm are summarized by Callahan (2008): while the LAN and Nref correlate to the bonding stage, N400 and P600 reflect to the resolution stage. However, as we can see from Table 7, instead of LAN (250 – 500 ms) and Nref (approximately 300 ms after stimulus onset), only the P200 effect (150 – 250 ms) is comparable to the LAN and Nref before the resolution stage (i.e., N400 and P600). Thus, the P200 is here taken as reflecting the bonding processing stage. The mapping of the two-stage model and the ERP findings during one antecedent anaphor resolution are made separately by antecedents' explicitness in each processing stage of the two-stage model in the three language groups.

First, the bonding stage of the two-stage model. Since no significant differences are reported between antecedent of definitional and stereotypical role nouns in natural gender language (English) and genderless language (Chinese), we may infer that at the bonding stage, the type of antecedent's gender (definitional or stereotypical, i.e., antecedent's gender explicitness) is not relevant for the processing of one antecedent anaphor resolution. As concluded by Osterhout, Bersick and McLaughlin (1997), as long as the gender is activated at the antecedent position (either definitional or stereotypical), it is ready for anaphoric pronoun resolution (Osterhout, Bersick & McLaughlin, 1997, in English; Su, Molinaro, Gillon-Dowens, Tsai, Wu & Carreiras, 2016, in Chinese). This inference probably may be generalized also to grammatical gender languages, because no significant findings preceded the reported N400 and P600 effects (Schmitt, Larmers & Münte, 2002). Importantly, it may also be possible that it is the linguistic context (i.e., anaphor resolution structure) that constrains the testability (or visibility) of antecedent's gender explicitness because the target words are the pronouns instead the antecedents.

P200 effects are reported during one antecedent anaphor resolution only in a genderless language (i.e., Mandarin Chinese). Due to the asymmetry of gender specificity encoded in Chinese third person pronouns, pronoun recognition and the search for the best-fit between the pronoun and its antecedent is clarified by the finding of the P200 in Experiment I (i.e., Su, Molinaro, Gillon-Dowens, Tsai, Wu & Carreiras, 2016). Thus, this component may reflect more the bonding stage of the two-stage model. By taking advantage of the unequal gender specificity encoded in Chinese third person pronouns, the connection between the bonding stage of the two-stage model and the P200 component can be better appreciated (see Table 8). In addition, it is interesting to note that when a pronoun is applicable to both genders, the processing cost for anaphoric resolution occurs mainly in the early stage and no cost for the later stage.

Second, the resolution stage of the two-stage model. The mapping of this resolution stage with available ERP findings differs by languages of distinct gender marking systems. As shown in Table 8, N400 and P600 effects are reported in Mandarin Chinese and German; no N400 but only P600 effects are reported in English. As summarized by Callahan (2008), both N400 and P600 effects can reflect

the processing of information integration during anaphor resolution. While N400 effects reflect the ease of semantic processing driven by antecedent and/or pronoun, P600 reflect integration difficulties resulting from the antecedent and the P600 integration difficulties can be either semantic or syntactic or both. Accordingly, based on Callahan (2008), two subordinate phases may be further differentiated in this resolution stage, i.e., an initial phase subject to possible interference and the wrap-up integration phase.

In the initial “interference phase”, interference can be driven by information encoded in the antecedent or the pronoun itself. For instance, antecedent’s gender explicitness can cause such interference in a grammatical gender language, i.e., German. Schmitt, Lamers & Münte (2002) reported N400 effects in sentences with antecedents of non-diminutive definitional role nouns as compared with diminutive ones. Similar semantics-related N400 effects are also found in Irmen, Holt & Weisbrod (2010) for male-biased antecedent as opposed to female-biased antecedents. A pronoun’s gender specificity can also lead to similar effects in a genderless language (i.e., Mandarin Chinese, Experiment I, a N400 effect is found for gender specific 她自己/*herself* as compared with 他自己/*himself*). Taken together, the interferences respectively reported in Mandarin Chinese and in German might basically reflect the processing cost due to formal interference probably driven by an expected word form (e.g., the general pronoun, anaphoric noun phrases or morpho-syntactic gender specific pronoun) (for similar N400 effects related to the expected word forms, please see Wicha, Moreno & Kutas, 2004; Barber & Carreiras, 2005; DeLong, Urbach & Kutas, 2005; Lau, Phillips, & Poeppel, 2008; Molinaro, Conrad, Barber, & Carreiras, 2010). Thus, at the initial interference phase, the factor that gives rise to the interference differed by gender marking systems encoded in the two languages. While the intervention is caused by gender specificity of Chinese third person pronouns, that in German is driven by antecedent’s gender explicitness (see Table 8).

As for the wrap-up integration phase, it may reflect the final wrap-up of gender information, and thus is the last step for the resolution stage. The P600 effects found for gender mismatch pronouns reported in all the studies shown in Table 7 are evidence for this wrap-up. Processing feature of this final integration differs depending on each tested gender marking system: semantics-based in Mandarin Chinese and German (when stereotypical gender and anaphora are manipulated in a grammatically correct sentence), syntactic constrained in English (i.e., syntactic gender agreement between antecedent and pronoun) and grammatical in German (i.e., grammatical gender agreement between antecedent gender and pronoun). Since the integration in this phase proceeds with the consequence resulting from the previous interference, independent gender integration system is the key in this wrap-up integration phase (e.g., semantic gender consistency or morpho-syntactic gender agreement between the antecedent and pronoun).

A special case can be spotted in this overall story. The semantics-based integration difficulty P600 effect found in Irmen et al. (2010) may be driven by the use of inclusive plural noun phrases (i.e., these men, these women or the people) for the purpose to avoid morpho-syntactic interference during one antecedent anaphor resolution. Given the strong effect of default stereotypical gender elicited by stereotyped role nouns reported in Experiment II (Chapter Three) and Experiment III (Chapter Four), if the anaphora was replaced by inclusive noun phrases, would similar semantics-based P600 effects be expected in sentences like (14) respectively in Chinese and English? Would the ERP findings differ by different gender markings? In

this case, the type of anaphora (singular third person pronouns or inclusive noun phrases) turns to be critical during one antecedent anaphor resolution. These questions should be answered in future studies.

(14) In Chinese: 廣場上，礦工們聚集在一起要求加薪，這些男人們/女人們開始對著政府官員吼叫。

In English: In the plaza, the miners gathered together asking for higher pay, these men/\*these women started to yell at the governor.

**Table 8. The inferred mapping of ERP Components and the two-stage model in each language group**

Language Group	Bonding Stage	Resolution Stage	
		initial interference phase	wrap-up integration phase
genderless language (Mandarin Chinese)	definitional and stereotypical P200 (perceptual)	N400 (pronoun's gender specificity: general > specific)	P600 (semantic)
natural gender language (English)	definitional and stereotypical		P600 (syntactic)
grammatical gender language (German)	definitional and stereotypical	N400 (definitional: diminutive > non-diminutive; stereotypical: female-biased > male-biased)	P600 (definitional: grammatical; stereotypical: semantic)

### 5.3 The Processing of Stereotypical Gender in Noun Phrases across Different Gender Systems

In the ERP literature, most studies related to manipulations of gender information usually focused on non-animate nouns phrases (Molinaro, Barber & Carreiras, 2011; Hagoort and Brown, 1999; Wicha, Bates, Moreno & Kutas, 2003; Wicha, Moreno & Kutas, 2003; Gunter, Friederici & Schriefers, 2000; Barber & Carreiras, 2005; Molinaro, Vespignani & Job, 2008; O'Rourke & Van Petten, 2011; Caffarra, Siyanova-Chanturia, Pesciarelli, Vespignani & Cacciari, 2015) and the foci were related to how grammatical or semantic information influences language processing. Sometimes grammatical gender was manipulated on animate noun phrases (Deutsch & Bentin, 2001; Silva-Pereyra, Gutierrez-Sigut & Carreiras, 2012). However, when and how the semantic animate/non-animate cues interacted with the syntactic information is the focus in these studies.

Up to now, Experiment II (in Spanish, i.e., Molinaro, Su & Carreiras, 2016) and Experiment III (in Chinese) are the only two ERP studies directly measuring the online activation of stereotypical knowledge on the target role nouns. Based on the available results, it seems that the immediacy of stereotypical knowledge is similar across languages, i.e., a default stereotypical gender is activated as long as a role noun is encountered. However, the processing mechanism and the linguistic representation of stereotypical knowledge indeed may be language specific in accordance with each distinct gender marking system.

In Experiment II (carried out in Spanish, a grammatical gender language), stereotypical gender was found to dominate the processing of stereotypical noun

phrases. In fact, N400 mismatch effects were found for all the three manipulated stereotypical gender mismatches in the noun phrases (determiner + role noun). Thus, stereotypical knowledge emerged as lexically represented in the word stem, for example, *miner-* (miner in English), male biased. This conclusion is consistent with Carreiras, Garnham and Oakhill (1996) who showed that the activation of stereotypical gender is “[...] *not necessary for local cohesion*” (Carreiras, Garnham, Oakhill & Cain, 1996, p. 659). It may be activated through lexical access of role nouns regardless of the grammar encoded in the text and thus against the minimalist hypothesis (proposed by McKoon & Ratcliff, 1992) (See chapter 1, section 2.1, p. 26).

On the other hand, in Experiment III (carried out in Mandarin Chinese, a genderless language), a N400 semantic mismatch effect was found only in the incongruent condition as compared with the default and no differences were found in other two comparisons (congruent vs. incongruent and congruent vs. default). Thus, according to the results, the activation of stereotypical gender actually can be modulated to some extent by the presence of gender modifiers before a role noun (i.e., the comparison between congruent vs. incongruent during the N400 time interval). This finding is consistent with the claim that the activation of stereotypical gender on the role nouns can be overridden if the gender is available before the target role noun is encountered (see cataphoras: Kreiner, Sturt & Garrod, 2008). By comparing the activation of definitional gender and stereotypical gender in anaphoric and cataphoric pronoun resolution structures, Kreiner, Sturt and Garrod (2008) concluded that while definitional gender is lexically represented, stereotypical gender is inferential and is processed when necessary, meaning that it is context-dependent. The processing of stereotypical gender in Chinese noun phrases is in line with these authors’ claim but here the gender is activated due to the linguistic context of a noun phrase instead of the linguistic structure of cataphoric pronoun resolution. Even so, stereotypical knowledge could not be overwritten by semantic gender even in Mandarin Chinese because a post-semantic negativity was found for the incongruent condition (as opposed to the congruent and the default during 800 ms – 1000 ms post stimulus). We suggest that this post-semantic negativity resembled the long-lasting negativity reported for stereotypical gender mismatches in the Spanish experiment (i.e., the N400 effects reported in Experiment II, Chapter Three) and both results give evidence to the special status of stereotypical knowledge that is different from general semantic knowledge.

Based on the results reported in Experiment II and III, we concluded that while stereotypical knowledge is found to be lexically represented in Spanish word stems (a grammatical gender languages), that in Chinese (a genderless language) is inferential and context-dependent. Such difference demonstrates that the mechanisms for the processing of stereotypical knowledge are affected by the gender marking system of the testing language. If the processing of stereotypical knowledge is language specific, we wonder how stereotypical knowledge is represented in English role nouns (a natural gender language) because role nouns in English can be categorized into two types: similar to Spanish, some role nouns have gender defined word-stem in the words (e.g., policeman/policewoman, chairman/chairwoman, actor/actress ... etc.) and, similar to Chinese, other role nouns that should have gender modifiers ahead to clearly express its gender (e.g., male/female babysitter, male/female doctor, male/female receptionist, ... etc.). In English, would the processing mechanism of stereotypical knowledge be diverse based on these two noun categories? And most importantly, if the N400 semantics-related effect could also be found in stereotypical gender incongruent noun phrases in English, could the special status of stereotypical



knowledge (i.e., stereotypical knowledge over semantic knowledge) be further confirmed beyond any gender marking systems across different languages?

## 5.4 Future Directions on the Processing of Gender Information

Behind the unanswered questions opened in this Discussion, further directions of research can be sketched.

For the processing of gender information, some studies focused on animate role nouns and some studies focused on non-animate object nouns. Among studies on animate role nouns, human role nouns have been studied the most. Given that gender is also an important biological feature in animals, it would also be interesting to study non-human nouns (i.e., animal nouns) to see how gender information is processed in animal nouns. It is clear that the process of human role nouns (animate) is different from that of object nouns (non-animate), but it remains unclear whether there is any distinction between human and non-human living things. It may be worth investigating the processing of gender information in animal nouns because this is related to how gender information is stored in the mental lexicon: animate vs. non-animate or human vs. non-human.

A series of three experiments could be developed. First, investigating if the processing of animal nouns is similar to that of object nouns, i.e., be treated as non-living things as objects. Thus, by using the ERP mismatch paradigm, P600 effects or LAN plus P600 effects might be expected in grammatical gender languages when morpho-syntactic gender disagrees within the noun phrase (e.g., *el perro*, the dog in Spanish, contrasted with *\*la perro*) (pure syntactic in grammatical gender languages, Hagoort and Brown, 1999; Wicha, Bates, Moreno & Kutas, 2003; Wicha, Moreno & Kutas, 2003; Gunter, Friederici & Schriefers, 2000; Barber & Carreiras, 2005; Molinaro, Vespignani & Job, 2008; O'Rourke & Van Petten, 2011; Caffarra, Siyanova-Chanturia, Pesciarelli, Vespignani & Cacciari, 2015). This grammatical gender interference could only be tested on animal nouns in grammatical gender languages and the control is noun phrases referring to objects. Here, it might also be possible that the semantic meaning leads the processing when the noun phrase is encountered and results in similar N400 effect as reported in Chapter Two (i.e., Molinaro, Su & Carreiras, 2016).

Second, testing if the processing of animal nouns is similar to that of definitional role nouns for the fix gender presented in the word. For instance, the gender for *el perro* (the dog) is masculine and this could be tested via anaphoric pronoun resolution structure (e.g. in Spanish: *El perro está sin hogar, él/ella pide comida a lo largo de la calle*; in English: *The dog is homeless, he/she/it asks for food along the street.*). Similar structure could be employed across the three language groups. How is the gender of a dog processed in languages without grammatical gender encoded in animal nouns, such as English (natural gender language) or Chinese (genderless language)? If each animal noun has its own default gender like definitional role nouns, similar results (as found in sentences with definitional antecedents as listed in Table 7) may be expected during one antecedent anaphor resolution. Even pronouns representing neutral gender could be tested (possibly could also be no gender or non-animate, e.g., *it* in English or 它/*ta*/ in Mandarin Chinese). In this case, would gender representation of animal noun differ by the arbitrary gender (e.g., while the mouse in Spanish is masculine as “*el ratón*” and it in German is feminine as “*die maus*”) encoded in a language? In support, Boroditsky and Schmidt (2000) claimed that

grammatical gender could be transformed into meaningful semantic gender, thus differing in languages of distinct gender marking systems?

Third, seeing if the processing of animal nouns may be similar to that of stereotypical role nouns. Although each animal noun can refer to both the male kind and the female kind, it may be biased by its characteristics. For instance, the lion may be male-biased for its role as a fierce predator and the rabbit may be female-biased for its role as an innocent victim. If this is the case, similar results (as found in sentences with stereotypical antecedents as listed in Table 7) may be expected during one antecedent anaphor resolution. Thus, could we say that the processing for social knowledge or stereotypical knowledge is different from other semantic meanings (Contreras, Banaji & Mitchell, 2012), is not unique to human, but to all animate living things?

On a different point of view, behind gender processing in animal nouns, neurocomputational models related to the processing of gender information could also be an issue if distinct gender marking systems are taken into account. Throughout the present dissertation and the studies listed in Table 7, three ERP components (P200, N400 and P600) are reported to index the processing of animate (specifically human only) gender information across languages of distinct gender marking systems. While P600 effects are reported as reflecting overall integration difficulties (semantic and/or syntactic or information updating) during sentence processing, N400 effects are reported as mirroring semantic anomalies related to pronoun's gender specificity, role noun's gender explicitness and role noun's stereotypical gender inconsistency (see Table 9). The functional representation of these two components corresponds to the neurocomputational model of general language processing presented by Brouwer and colleagues (2012; in press): the Retrieval-Integration (RI) account. According to Brouwer and colleagues (2012; in press), "the N400 component reflects retrieval of lexical-semantic information, rather than semantic integration or any other kind of compositional semantic processing" (Brouwer and colleagues, in press, p.4) and the P600 component reflects "difficulties in establishing a coherent utterance representation, rather than processes operating on a purely syntactic representation" (Brouwer and colleagues, in press, p.5).

Thus, if N400 effects could also be found during the processing of stereotypical role nouns in natural gender languages (e.g., English), the special status of stereotypical gender could be generally viewed as a special semantic category beyond general semantic knowledge. If this is the case, the computational model of the Retrieval-Integration (RI) account may support the processing of gender information in languages belonging to natural gender languages and grammatical gender languages (see Table 9). While the N400 effect could represent the retrieval of lexical-semantic information, the P600 effect may reflect the establishment of a coherent utterance representation.

However, when a differential language system is considered (e.g., Mandarin Chinese, a genderless language as presented in Chapter Two and Chapter Four in the present dissertation), the Retrieval-Integration (RI) account may be extended as Perceptual-Retrieval-Integration (PRI) account for the perceptual-related P200 component generally found in Mandarin Chinese (Liu et al., 2003; Lee et al., 2007; Hsu et al., 2009; Yum et al., 2014; Hung and Schumacher, 2012, 2014). Further empirical evidence is necessary for this proposal.

**Table 9. The summary of the ERP components found in the present dissertation and some previous representative studies related to the processing of gender information**

language group	lexical-related ERP effects on the target words	congruity-related ERP effects	
		anaphoric pronoun resolution (unambiguous mismatch between the antecedent and the target pronoun)	stereotyped noun phrases (gender mismatch between the modifier/determiner/role noun's gender morpheme and the default stereotypical gender of the target role noun)
genderless languages	P200 and N400 effects on the gender specificity of pronouns	P200 and P600 mismatch effects	N400 and late negativity (800 ms - 1000 ms) mismatch effects
natural gender languages	?	P600 mismatch effects	?
grammatical gender languages	N400 effect for gender stereotypicality or semantic gender of a non-diminutive definitional role noun	P600 mismatch effects	N400 mismatch effects

## 5.5 Conclusion

The processing of human gender information (definitional gender and/or stereotypical gender) in languages of distinct gender marking systems is studied in the present thesis. The criteria that classify languages into different categories are by means of the presence/absence of morpho-syntactic gender morphemes and grammatical gender agreement required between a role noun and its related words. By using the ERP mismatch paradigm, the results confirmed: gender information is differently processed in accordance with the tested gender marking system of a certain language and the target words set in the linguistic frameworks (anaphora in anaphoric resolution or role nouns in noun phrases). In a genderless language (Mandarin Chinese) where gender specificity is differently encoded in the third person pronouns (他/*he* for every individual and 她/*she* for female only), P600 effects are found only in unambiguous gender mismatched conditions during anaphoric pronoun resolution (e.g., the Dalai Lama – 她/*she*). Compared to the explicit definitional gender, the implicit stereotypical gender is less visible during anaphoric pronoun resolution. When the observation is directly focused on the role nouns, the immediacy hypothesis (the default stereotypical knowledge is activated via the encounter of a role noun) is confirmed; but the processing mechanisms differed by languages of different gender markings. While stereotypical knowledge is found to lead the whole processing of Spanish noun phrases (regardless of the grammatically/stereotypically mismatched gender morphemes in the determiner and/or the role noun), Chinese gender markers can modulate the activation of stereotypical gender to some extent; but stereotypical gender overrode at a relatively late stage as opposed to that found in Spanish. The results presented in the present thesis suggested: stereotypical knowledge in Spanish is activated via lexical access of a word, and that in Chinese, is inferential and context dependent. Even so, stereotypical knowledge is confirmed as being different from general semantic knowledge in a grammatical gender language (i.e., Spanish) and a genderless language (i.e., Mandarin Chinese).

If the proposed further studies could be carried out in the future, we will surely have a better understanding on the processing of gender information in languages of distinct gender marking systems, not only to have empirical evidence to evaluate differences between different languages, but also to support the theoretical inferences or clarifications.

## References

- Banaji, M.R., and Hardin, C.D. (1996). Automatic stereotyping. *Psychological Science*, 7(3), 136-141.
- Banaji, M.R., Hardin, C.D., and Rothman, A.J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, 65(2), 272-281.
- Barber, H., and Carreiras, M. (2003). Integrating gender and number information in Spanish Word pairs: An ERP study. *Cortex*, 39, 465-482.
- Barber, H., and Carreiras, M. (2005). Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. *Journal of Cognitive Neuroscience*, 17(1), 137-153.
- Barber, H. A., Salillas, E., and Carreiras, M. (2004). Gender or genders agreement? In M. Carreiras & C. Clifton (Eds.), *On-line Study of Sentence Comprehension: Eye-tracking, ERP and Beyond* (pp. 309-328). Brighton, UK: Psychology Press.
- Barber, H., Vergara, M., and Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: Evidence from ERPs. *Neuroreport*, 15(3), 545-548.
- Blanchet, S., Gagnon, G., and Bastien, C. (2007). Event-related potential of dynamic neural mechanisms of semantic organizational strategies in verbal learning. *Brain Research*, 1170, 59-70.
- Boroditsky, L., and Schmidt, L. (2000). Sex, syntax and semantics. *Proceedings of the Cognitive Science Society*, 22, 42-47.
- Brouwer, H., Crocker, M.W., Venhuizen, N.J., and Hoeks, J.C.J. (in press). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive Science*, 1-35.
- Brouwer, H., Fitz, H., and Hoeks, J.C.J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127-143.
- Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98, 159-168.
- Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Cognitive Neuroscience and Neuropsychology*, 18(17), 1851-1854.
- Burmester, J., Spalek, K., and Wartenburger, I. (2014). Context updating during sentence comprehension: The effect of aboutness topic. *Brain and Language*, 137, 62-76
- Cacciari, C., and Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics*, 28(2), 277-293.
- Caffarra, S., Siyanova-Chanturia, A., Pesciarelli, F., Vespignani, F., and Cacciari, C. (2015). Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian. *Psychophysiology*, 52, 1019-1030.
- Callahan, S.M. (2008). Processing anaphoric constructions: Insights from electrophysiological studies. *Journal of Neurolinguistics*, 21, 231-266.
- Canal, P., Garnham, A., and Oakhill, J. (2015). Beyond Gender Stereotypes in Language Comprehension: Self Sex-Role Descriptions Affect the Brain's Potentials Associated with Agreement Processing. *Frontiers in Psychology*, 6, 1953.
- Carreiras, M., Garnham, A., and Oakhill, J. (1993). The use of superficial and meaning-based representations in interpreting pronouns: evidence from Spanish. *European Journal of Cognitive Psychology*, 5, 93-116
- Carreiras, M., Garnham, A., Oakhill, J., and Cain, K. (1996). The use of stereotypical

- gender information in constructing a mental model: evidence from English and Spanish. *Quarterly Journal of Experimental Psychology*, 49, 639–663.
- Carreiras, M., Salillas, E., and Barber, H. (2004). Event related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, 20, 98-105.
- Carreiras, M., Vergara, M., and Barber, H. (2005). Early ERP effects of syllabic processing during visual word recognition. *Journal of Cognitive Neuroscience*, 17(11), 1803-1817.
- Chang, B.-M. (2007). 《“她”從哪裡來？現代女性指稱的源流考釋》(Where does “她” come from? An enquiry into the origin of the modern feminine pronoun), 《歷史月刊》(*Historical Monthly*), 237, 123-127.
- Cherng, R.-J., Chang, C.-L., and Chen, J.-Y. (2009). A new look at gender inequality in Chinese: A study of Chinese speakers’ perception of gender-based characters. *Sex Roles*, 61, 427-433.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Foris.
- Cohen, P. (2012, December 11). More Women Are Doctors and Lawyers Than Ever—but Progress Is Stalling. *The Atlantic*. Retrieved August 27, 2016, from <http://www.theatlantic.com/sexes/archive/2012/12/more-women-are-doctors-and-lawyers-than-ever-but-progress-is-stalling/266115/>
- Contreras, J.M., Banaji, M.R., and Mitchell, J.P. (2012). Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Social Cognitive and Affective Neuroscience*, 7, 764-770.
- Corbett, G.G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Dambacher, M., Kliegl, R., Hofmann, M., and Jacobs, A.M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, 1084, 89-103.
- DeLong, K.A., Urbach, T.P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117-1121.
- Dictionary of Chinese Character Variants. (2000). Retrieved April 30, 2015, from <http://dict.variants.moe.edu.tw/eng.htm>
- Ding, G., Peng, D., and Taft, M. (2004). The nature of the mental representation of radicals in Chinese: A priming study. *Journal of Experimental Psychology: Learning, memory and cognition*, 30(2), 530-539.
- Duffy, S.A., and Keir, J.A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory and Cognition*, 32(4), 551-559.
- Deutsch, A., and Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew. *Journal of Memory and Language*, 45, 200-224.
- Esaulova, Y., Reali, C., and von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: Eye movements in pronominal and noun phrase anaphor resolution. *Language, Cognition and Neuroscience*, 29(7), 781-803.
- Ettner, C. (2002). In Chinese, men and women are equal – or – women and men are equal? In Hellinger Marlis & Hadumod Bußmann (Eds.) *Gender Across Languages: The Linguistic Representation of Women and Men* (pp. 29-55). Amsterdam: John Benjamins Publishing Company.
- Farris, C.S. (1988). Gender and Grammar in Chinese – With implications for language universals. *Modern China*, 14(3), 277-308.

- Feldman, L.B., and Siok, W.W.T. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language*, 40, 559-576.
- Friederici, A.D. (2011). The brain basis of language processing: From structure to function. *Physiological Review*, 91, 1357-1392.
- Friederici, A.D., Hahne, A., and Saddy, D. (2002). Distinct Neurophysiological Patterns Reflecting Aspects of Syntactic Complexity and Syntactic Repair. *Journal of Psycholinguistic Research*, 31(1), 45-63.
- Garnham, A., Oakhill, J., Ehrlich, M.F., and Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. *Journal of Memory and Language*, 34(1), 41-62.
- Garnham, A., Oakhill, J., and Reynolds, D. (2002). Are inferences from stereotyped role names to character's gender made elaboratively? *Memory and Cognition*, 30(3), 439-446.
- Garrod, S., and Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, 42, 526-544.
- Garrod, S., and Sanford, A.J. (1994). "Resolving sentences in a discourse context: How discourse representation affects language understanding," in M. A. Gerbsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 675-698). San Diego, CA: Academic Press.
- Grover, A. (2015, October 29). The Good and Bad Statistics on Women in Medicine. *The Wall Street Journal*. Retrieved August 27, 2016, from <http://blogs.wsj.com/experts/2015/10/29/the-good-and-bad-statistics-on-women-in-medicine/>
- Gunter, T.C., Friederici, A.D., and Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, 12(4), 556-568.
- Hagoort, P., and Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, 28(6), 715-728.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combing syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6), 883-899.
- Hagoort, P. (2009). "Reflections on the neurobiology of syntax," in D. Bickerton, & E. Szathmáry (Eds.), *Biological foundations and origin of syntax* (pp. 279-296). Cambridge, MA: MIT Press.
- Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. 16 April, 2004, *Science*, 304, 438-441.
- Hahne, A., and Friederici, A.D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, 13(3), 339-356.
- Hellinger, M., and Bußmann, H. (2001). The linguistic representation of women and men. In M. Hellinger & H. Bußmann (Eds.), *Gender across languages* (Vol. 1, pp. 1-25). Philadelphia: John Benjamins Company.
- Ho, S.-H., Ng, T.-T., and Ng, W.-K. (2003). A "radical" approach to reading development in Chinese: The role of semantic radicals and phonetic radicals. *Journal of Literacy Research*, 35(3), 849-878.
- Hsu, C.-H., Tsai, J.-L., Lee, C.-Y., and Tzeng, O.J.L. (2009). Orthographic

- combinability and phonological consistency effects in reading Chinese phonograms: An event-related potential study. *Brain and Language*, 108(2), 56–66.
- Hua, X.-H. (2012). 《近三十年“她”字研究綜述》 (An integrated discussion on the research of ‘她’ nearly thirty years), 《現代語文：語文研究版》 (*Modern Chinese*), 8, 6-8.
- Huang, C.-H., and Luh, W.-M. (2012). Analysis of third person pronouns with gender semantic features in elementary school Mandarin textbooks. *Journal of Textbook Research*, 5(1), 85-113.
- Hung, Y.-C., and Schumacher, P.B. (2012). Topicality matters: position-specific demands on Chinese discourse processing. *Neuroscience Letters*, 511, 59-64.
- Hung, Y.-C., and Schumacher, P.B. (2014). Animacy matters: ERP evidence for the multi-dimensionality of topic-worthiness in Chinese. *Brain Research*, 1555, 36-47.
- Irmen, L., Holt, D.V., and Weisbrod, M. (2010). Effects of role typicality on processing person information in German: Evidence from an ERP study. *Brain Research*, 1353, 133-144.
- Irmen, L., and Roßberg, N. (2004). Gender Markedness of Language The Impact of Grammatical and Nonlinguistic Information on the Mental Representation of Person Information. *Journal of Language and Social Psychology*, 23(3), 272-307.
- Jäger, L., Engelmann, F., and Vasishth, S. (2015). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, 6, 617.
- Jiang, X., and Zhou, X. (2009). Processing different levels of syntactic hierarchy: An ERP study on Chinese. *Neuropsychologia*, 47, 1282-1293.
- Kaan, E., Harris, A., Gibson, E. and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2), 159-201.
- Kaan, E., and Swaab, T.Y. (2003). Repair, revision, complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15(1), 98-110.
- Kennison, S.M., and Trofe, J.L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, 32(3), 355-377.
- Kim, A., and Osterhout, L. (2005). The independence of combinatory semantic processing: evidence from event-related potentials. *Journal of Memory and Language*, 52(2), 205-225.
- Kolk, H.H.J., Chwilla, D.J., van Herten, M., and Oor, P. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain and Language*, 85, 1-36.
- Kreiner, H., Mohr, S., Kessler, K., and Garrod, S. (2009). Can context affect gender processing: Evidence from ERP about the differences between definitional and stereotypical gender. In K. Alter, M. Horne, M. Lindgren, M. Roll, & J. von Koss Torkildsen (Eds.), *Brain Talk: Discourse with and in the Brain* (pp. 107-119).
- Kreiner, H., Sturt, P., and Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58(2), 239-261.
- Kuperberg, G.R. (2007). Neural mechanisms of language comprehension: Challenges



- to syntax. *Brain Research*, 1146, 23-49.
- Kuperberg, G.R., Caplan, D., Sitnikova, T., Eddy, M., and Holcomb, P.J. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes*, 21(5), 489-530.
- Kutas, M., and Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the Event-related brain potential (ERP). *The Annual Review of Psychology*, 62(14), 1-27.
- Kutas, M., and Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.
- Kutas, M., and Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161-163.
- Lattner, S., and Friederici, A.D. (2003). Talker's voice and gender stereotype in human auditory sentence processing – evidence from event-related brain potentials. *Neuroscience Letters*, 339, 191–194.
- Lau, E.F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9, 920-933.
- Lee, C.-Y., Liu, Y.-N., and Tsai, J.-L. (2012). The time course of contextual effects on visual word recognition. *Frontiers in Psychology*, 3, 285.
- Lee, C.-Y., Tsai, J.-L., Chan, W.-H., Hsu, C.-H., Hung, D.L., and Tzeng, O.J.L. (2007). Temporal dynamics of the consistency effect in reading Chinese: An event-related potentials study. *Neuroreport*, (18)2, 147-151.
- Li, C.N., and Thompson, S.A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Li, Y.H.A., and Shi, Y.Z. (2000). 《漢語量詞系統的建立與複數標記“們”的發展》 (The story of “men”). *Contemporary Linguistics*, 2(1), 27–36.
- Ling, Y.-Z. (1989). 《“她”字的創造歷史》 (The history of the character ‘她’), 《語言教學與研究》 (*Language Teaching and Research*), 4, 139-151.
- Liu, Y., Perfetti, C.A., and Hart, L. (2003). ERP evidence for the time course of graphic, phonological, and semantic information in Chinese meaning and pronunciation decisions. *Journal of Experimental Psychology: Learning, memory and cognition*, 29, 1231-1247.
- Luck, S., and Hillyard, S. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31(3), 291–308.
- Martin-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., and Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun–adjective agreement in Spanish. *Brain Research*, 1093, 178-189.
- Molinaro, N., Barber, H.A., and Carreiras, M. (2011a). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 47, 908- 930.
- Molinaro, N., Barber, H.A., Caffarra, S., and Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement: A reply to Tanner et al.. *Cortex*, 66, 156-159.
- Molinaro, N., Carreiras, M., and Duñabeitia, J.A. (2012). Semantic combinatorial processing of non-anomalous expressions. *NeuroImage*, 59, 3488-3501.
- Molinaro, N., Conrad, M., Barber, H., and Carreiras, M. (2010). On the functional nature of the N400: Contrasting effects related to visual word recognition and contextual semantic integration. *Cognitive Neuroscience*, 1(1), 1-7.
- Molinaro, N., Kim, A., Vespignani, F., and Job, R. (2008a). Anaphoric agreement violation: An ERP analysis of its integration. *Cognition*, 106(2), 963-974.
- Molinaro, N., Su, J.-J., and Carreiras, M. (2016). Stereotypes override grammar:

- Social knowledge in sentence comprehension. *Brain and Language*, 155/156, 36-43.
- Molinaro, N., Vespignani, F., and Job, R., (2008b). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research*, 1228, 161-176.
- Molinaro, N., Vespignani, F., Zamparelli, R., and Job, R., (2011b). Why brother and sister are not just siblings? Repair processes in agreement computation. *Journal of Memory and Language*, 64(3), 211-232.
- Moser, D. (1997). Covert sexism in Mandarin Chinese. *Sino-Platonic Papers*, No. 74, January 1997, 1-23.
- Munte, T.F., Heinze, H.-J., Matzke, M., Wieringa, B.M., and Johannes, S. (1998). Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. *Neuropsychologia*, 36(3), 217-226.
- Nieuwland, M.S., and Van Berkum, J.J.A. (2006a). When peanuts fall in love: N400 Evidence for the power of Discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098-1111.
- Nieuwland, M.S., and Van Berkum, J.J.A. (2006b). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, 1118, 155-167.
- Norris, C.J., Chen, E.E., Zhu, D.C., Small, S.L., and Cacioppo, J.T. (2004). The interaction of social and emotional processes in the brain. *Journal of Cognitive Neuroscience*, 16(10), 1818-1829.
- Oakhill, J., Garnham, A., and Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory and Cognition*, 33(6), 972-983.
- O'Rourke, P.L., and Van Petten, C. (2011). Morphological agreement at a distance: Dissociation between early and late components of the event-related brain potential. *Brain Research*, 1392, 62-79.
- Osterhout, L., Bersick, M., and McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition*, 25(3), 273-285.
- Osterhout, L., and Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.
- Osterhout, L., and Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, 14, 283-317.
- Packard, J.L. (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach*. CAMBRIDGE UNIVERSITY PRESS.
- Peng, J.-J. (2009). 《新詞“她 X”探析》 (Discussion on the new word 'She X'), 《經濟研究導刊》 (*Economic Research Guide*), 1, 235-236.
- Perfetti, C.A., and Zhang, S. (1991). Phonetic processes in reading Chinese words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17(4), 633-643.
- Pyykkönen, P., Hyönä J., and van Gompel, R.P.G. (2010). Activating gender stereotypes during online spoken language processing. *Experimental Psychology*, 57(2), 126-133.
- Qiu, L., Swaab, T.Y., Chen, H-C, and Wang, S. (2012). The role of gender information in pronoun resolution: Evidence from Chinese. *PLoS ONE*, 7(5), e36156.
- Rabovsky, M., and McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, 132, 68-89.

- Reynolds, D.J., Garnham, A, and Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *The Quarterly Journal of Experimental Psychology*, 59(5), 886-903.
- Sabbagh, M.A., Moulson, M.C., and Harkness, K.L. (2004). Neural correlates of mental state decoding in human adults: An event-related potential study. *Journal of cognitive neuroscience*, 16(3), 415-426.
- Sabourin, L., and Stowe, L.A. (2008). Second language processing: When are first and second languages processed similarly. *Second Language Research*, 24(3), 397-430.
- Schmitt, B.M., Lamers, M., and Münte, T.F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cognitive Brain Research*. 14(3), 333-346.
- Siewierska, A. (2013). "Gender Distinctions in Independent Personal Pronouns," in M.S. Dryer & M. Haspelmath, Martin (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved August 06, 2015 from <http://wals.info/chapter/44>
- Silva-Pereyra, J., Gutierrez-Sigut, E., and Carreiras, M. (2012). An ERP Study of Coreference in Spanish: Semantic and Grammatical Gender Cues. *Psychophysiology*, 49, 1401-1411.
- Siyanova-Chanturia, A., Pesciarelli, F., and Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS ONE*, 7(12), e48712.
- Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). "Representation of the sexes in language". In K. Fiedler (Ed.), *Social Communication* (pp. 163–187). New York: Psychology.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542-562.
- Su, J.-J., Molinaro, N., Tsai, P.-S., Wu, D.H., and Carreiras, M. (submitted). When "he" can also be "she": An ERP study of reflexive pronoun resolution in Mandarin Chinese, *Frontiers in Psychology*, 7, 151.
- Tang, T.-C. (1988). 《國語詞彙的『重男輕女』現象》 (Male chauvinism in the Chinese lexical structure), in Tang, T.-C. (Ed.), 《漢語詞法句法論集》 (*Studies on Chinese Morphology and Syntax*) (pp. 59-65). Taipei: Student Book Co., Ltd.
- Tsai, P.-S., Yu, B. H.-Y., Lee, C.-Y., Tzeng, O.J.L., Hung, D.L., and Wu, D.H. (2009). An event-related potential study of the concreteness effect between Chinese nouns and verbs. *Brain Research*, 1253, 149-160.
- Tsang, Y.-K., and Chen, H.-C. (2009). Do position-general radicals have a role to play in processing Chinese characters? *Language and Cognitive Processes*, 24 (7/8), 947-966.
- Van Berkum, J.J.A., Van den Brink, D., Tesink, C.M.J.Y., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580–591.
- Van de Meerendonk, N., Kolk, H.H.J., Chwilla, D.J., and Vissers, C.Th.W.M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, 3(5), 1211-1224.
- Van de Meerendonk, N., Kolk, H.H.J., Vissers, C.W.M., and Chwilla, D.J. (2010). Monitoring in Language Perception: Mild and Strong Conflicts Elicit Different ERP Patterns. *Journal of Cognitive Neuroscience*, 22(1), 67-82.
- Van den Brink, D., Van Berkum, J.J.A., Bastiaansen, M.C.M., Tesink, C.M.J.Y., Kos,

- M., Buitelaar, J.K., and Hagoort, P. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social, Cognitive and Affective Neuroscience*, 7, 173-183.
- Van Herten, M., Kolk, H.H.J., and Chwilla, D.J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241-255.
- Van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, 18(4), 380-393.
- Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., and McIsaac, H. (1991). Fractionating the word repetition effect with Event-Related Potentials. *Journal of Cognitive Neuroscience*, 3(2), 131-150.
- Varnum, M.E.W., Na, J., Murata, A., and Kitayama, S. (2012). Social Class Differences in N400 indicate differences in spontaneous trait inference. *Journal of Experimental Psychology: General*, 14(3), 518-526.
- White, K.R., Crites Jr., S.L., Taylor, J.H., and Corral, G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Social Cognitive and Affective Neuroscience*, 4, 191-198.
- Wang, L., and Schumacher, P.B. (2013). New is not always costly: evidence from online processing of topic and contrast in Japanese. *Frontiers in Psychology*, 4, 363
- Wang, P., Yang, Y.P., and Zhao, L. (2010). The activation of stereotypes: Behavioral and ERPs evidence. *Acta Psychologica Sinica*, 42(5), 607-617.
- Wang, Y.-Y. (2010). 《從‘ta’看‘他’與‘她’》(From ‘ta’ to look at ‘他’ and ‘她’), 《大眾文藝》(*Da Zhong Wen Yi*), 6, 78.
- Wicha, N.Y.Y., Moreno, E.M., and Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, 39, 483-508.
- Wicha, N.Y.Y., Bates, E.A., Moreno, E.M., and Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346, 165-168.
- Wicha, N.Y.Y., Moreno, E.M., and Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16(7): 1272-1288.
- Word List with Accumulated Word Frequency in Sinica Corpus. (2005). Retrieved April 30, 2015 from [http://elearning.ling.sinica.edu.tw/eng\\_teaching\\_index.html](http://elearning.ling.sinica.edu.tw/eng_teaching_index.html)
- Wu, D.-S., and Liang, T. (2008). Chinese pronominal anaphora resolution using lexical knowledge and entropy-based weight. *Journal of the American Society for Information Science and Technology*, 59(13). 2138-2145.
- Xu, X.-D., Jiang, X.-M., and Zhou, X.-L. (2013). Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation. *Brain and Cognition*, 81, 223-236.
- Ye, Z., Zhan, W., and Zhou, X. (2007). The semantic processing of syntactic structure in sentence comprehension: An ERP study. *Brain Research*, 1142, 135-145.
- Yum, Y.-N., Law, S.-P., Su, I.-F., Lau, D. K.-Y., and Mo, K.-N. (2014). An ERP study of effects of regularity and consistency in delayed naming and lexicality judgment in a logographic writing system. *Frontiers in Psychology*, 5, 315.

## Appendix A. Proper names (definitional role nouns) used in Experiment I

	item	Chinese	Translation into English (based on the celebrity's personal website or wiki, if not the Pinyin is used)		item	Chinese	Translation into English (based on the celebrity's personal website or wiki, if not the Pinyin is used)
male proper names	1	言承旭	Jerry Yan	female proper names	1	林慧萍	Monique Lin
	2	李登輝	Teng-Hui Lee		2	劉嘉玲	Carina Lau
	3	魏德聖	Te-Sheng Wei		3	張鈞甯	Janine Chang
	4	阿基師	Chef Aki		4	劉真	Serina Liu
	5	林義傑	Kevin Lin		5	蔡英文	Ing-Wen Tsai
	6	羅斯福總統	President Roosevelt		6	侯佩岑	Patty Hou
	7	蕭敬騰	Jam Hsiao		7	呂秀蓮	Annette Lu
	8	王建民	Chien-Ming Wang		8	李毓芬	Tia Lee
	9	張菲	Fei Chang		9	陳菊	Kiku Chen
	10	達賴喇嘛	Dalai Lama		10	桂綸鎂	Lun-Mei Gui
	11	馬英九	Ying-Jeou Ma		11	林青霞	Brigitte Lin
	12	劉德華	Andy Lau		12	酒井法子	Noriko Sakai
	13	吳宗憲	Jacky Wu		13	莫文蔚	Karen Mok
	14	高凌風	Frankie Kao		14	安心亞	Amber An
	15	彭于晏	Eddie Yu-Yen Peng		15	林熙蕾	Kelly Lin Hsi-Lei
	16	周杰倫	Jay Chou		16	Selina	Selina
	17	楊志良	Zhi-Liang Yang		17	林嘉綺	Patina Lin
	18	王力宏	Lee Hong Wang		18	楊丞琳	Rainie Yang
	19	豬哥亮	Yu-Chen Hsieh		19	徐若瑄	Vivian Hsu
	20	李安	Ang Lee		20	舒淇	Qi Shu
	21	庾澄慶	Harlem Yu		21	林志玲	Chiling Lin
	22	謝霆鋒	Nicholas Tse		22	周美青	Christine Chow
	23	黑人陳建州	Blackie		23	張柏芝	Cecilia Cheung
	24	蔡康永	Kevin Tsai		24	蔡依林	Jolin Tsai
	25	周華健	Wa Kin Chau		25	小S	Dee
	26	楊宗緯	Aska Yang		26	范瑋琪(范范)	Christine Fan
	27	吳伯雄	Bo Xiong Wu		27	蕭淑慎	Suzanne Hsiao
	28	康康	Kan Kan		28	于美人	Belle Yu
	29	張學友	Jacky Cheung		29	周迅	Jue
	30	郭台銘	Terry Guo		30	趙薇	Vicki Zhao
	31	金城武	Kaneshiro Takeshi		31	溫嵐	Landy
	32	宋楚瑜	James C.Y. Soong		32	陶晶瑩	Matilda Tao
	33	李敖	Ao Li		33	劉若英	Rene Liu
	34	金賢重	Kim Hyun Joong		34	范冰冰	Bing Bing Fan
	35	陳水扁	Shui-bian Chen		35	伊能靜	Annie Yi
	36	大炳	Tony Fish		36	關芝琳	Rosamund Kwan Chi-lam
	37	陳昇	Bobby Chen		37	邱淑貞	Chingmy Yau
	38	柯震東	Kai Ko		38	林依晨	Ariel Lin
	39	李李仁	Lee-Zen Lee		39	梁洛施	Isabella Leong Lok-sze
	40	齊秦	Chin Chyi		40	大S	Barbie Hsu

## Appendix B. Stereotypical role nouns used in Experiment I

	item	Chinese	Translation into English	Mean male-biased score
male biased	1	搬運工人	porter	91.05
	2	將軍	general	90.71
	3	礦工	miner	90.53
	4	陸軍上校	colonel	90.00
	5	強姦犯	rapist	89.82
	6	棒球選手	baseball player	89.65
	7	海軍上將	admiral	89.30
	8	貨櫃車司機	lorry driver	88.95
	9	海軍上尉	lieutenant	88.42
	10	拳擊手	boxer	88.25
	11	水電工	plumber	88.07
	12	焊接工人	welder	87.89
	13	馬路工人	road worker	87.72
	14	保全	security guard	87.19
	15	重型機械操作員	heavy equipment operator	87.19
	16	連續殺人魔	serial killer	87.02
	17	消防員	fire fighter	86.43
	18	建築工人	construction worker	86.32
	19	足球教練	football coach	85.96
	20	警長	police sergeant	85.79
	21	警衛	guard	85.71
	22	足球員	football player	85.61
	23	傘兵	parachutist	85.44
	24	修屋頂的人	roofer	85.09
	25	士兵	soldier	84.91
	26	泥水匠	bricklayer	84.74
	27	船員	sailor	84.74
	28	武術指導	martial art instructor	84.39
	29	修車技師	car mechanic	84.21
	30	獵人	hunter	84.04
	31	典獄長	warden	83.86
	32	賽車手	racing car driver	83.51
	33	總統	president	83.33
	34	監獄看守員	jailor	82.81
	35	乩童	a person who speaks for the god	82.46
	36	建築承包商	building contractor	82.28
	37	木匠	carpenter	82.11
	38	太空人	astronaut	81.40
	39	列車長	train conductor	81.23
	40	木工	wood worker	81.23

	item	Chinese	Translation into English	Mean male-biased score
female biased	1	粉絲	fan	32.98
	2	美髮師	hairstylist	32.28
	3	食堂打菜人員	canteen assistant	31.93
	4	塔羅牌專家	tarot card reader	31.93
	5	電話接線員	telephone operator	31.75
	6	舞蹈老師	dance instructor	31.40
	7	購物者	shopper	31.05
	8	花店業者	florist	30.00
	9	追星族	groupie	29.82
	10	彩妝師	make up artist	29.47
	11	小學老師	primary school teacher	29.12
	12	婚禮顧問	wedding planner	28.25
	13	時裝模特兒	fashion model	27.72
	14	長笛演奏家	flute player	27.02
	15	啦啦隊長	cheerleader	26.67
	16	社工人員	social worker	26.32
	17	模特兒	model	25.79
	18	手工編織師	weaver	25.54
	19	巧克力愛好者	chocolate lover	25.09
	20	美容師	beautician	23.68
	21	脫衣舞者	exotic dancer	23.68
	22	空服員	flight attendant	23.51
	23	賣淫者	prostitute	22.46
	24	護理人員	careworker	22.11
	25	電話總機人員	switchboard operator	21.75
	26	魔術師助理	magician's assistant	21.40
	27	芭蕾舞者	ballet dancer	20.70
	28	秘書	secretary	20.18
	29	幼兒保育員	childminder	19.65
	30	幼兒教師	infant teacher	19.30
	31	看護	caregiver	18.07
	32	看顧小孩者	babysitter	18.07
	33	媒人	matches man and woman for them	17.19
	34	托兒所教師	nursery teacher	16.79
	35	護士	nurse	16.67
	36	水中芭蕾舞者	synchronized swimmer	16.14
	37	褓姆	nanny	15.96
	38	被強姦者	rape victim	15.26
	39	幼稚園老師	kindergarten teacher	14.91
	40	美甲師	manicurist	13.33

**Appendix C. Male-biased rating scores for the 80 Spanish role nouns employed in the present experiment.**

The first forty role nouns are the most male-biased ones and the last forty the most female-biased ones. Log frequency is extracted by the EsPal data-base.

(<http://www.bcbl.eu/databases/espal/>).

Sequence no.	Spanish role noun	English translation	Male-biased rating score	Standard error	LogFreq masculine	LogFreq feminine
s01	violadores	male rapists	92.540	1.738	0.278757	0.009767
s02	mineros	male miners	91.429	1.058	0.833114	0.882056
s03	obreros	male labors	88.438	1.412	1.353053	1.071650
s04	marineros	male sailors	87.031	1.254	0.835182	0.463968
s05	camioneros	male road builders	86.875	1.065	0.449174	0.026016
s06	fontaneros	male plumbers	85.469	1.407	0.207053	0.005608
s07	bomberos	male fire fighters	85.000	1.179	0.388434	0.001823
s08	cazadores	male hunters	84.375	1.334	1.052196	0.513838
s09	carpinteros	male carpenters	82.656	1.359	0.721492	0.029985
s10	asesinos	male murderers	80.318	1.600	1.303094	0.757227
s11	carceleros	male jailers	79.524	1.899	0.458602	0.037815
s12	luchadores	male fighters	79.219	1.814	0.931027	0.249577
s13	basureros	male garbage collectors	77.656	1.913	0.255102	0.000431
s14	presidentes	male presidents	77.656	1.707	2.960595	2.007741
s15	gobernadores	male governors	76.719	1.682	1.859681	0.419597
s16	conquistadores	male conquerors	75.873	2.485	0.943901	0.142983
s17	magos	male magicians	73.125	1.650	1.069851	0.288306
s18	mayordomos	male butlers	72.540	2.771	0.850380	0.035221
s19	ladrones	male thieves	72.344	1.764	1.187986	0.255885
s20	prisioneros	male prisoners	72.344	1.602	1.170999	0.491670
s21	vagabundos	male vagabonds	72.222	1.559	0.589055	0.118981
s22	banqueros	male bankers	70.794	1.815	0.809918	0.020667
s23	carniceros	male butchers	70.781	1.758	0.556613	0.044235
s24	granjeros	male farmers	70.625	1.835	0.449174	0.045507
s25	presos	male prisoners	70.318	1.767	1.195427	1.488705
s26	entrenadores	male coaches	70.156	2.011	1.669703	0.184581
s27	alcohólicos	male alcoholics	70.000	1.494	0.454657	0.524933
s28	zapateros	male shoemakers	69.844	2.210	1.649781	0.061720
s29	payasos	male clowns	69.844	1.551	0.682552	0.020667
s30	cirujanos	male surgeons	69.531	1.613	0.852169	0.066589
s31	sospechosos	male suspects	69.365	1.905	0.966580	0.557395
s32	jefes	male chiefs	69.063	1.570	2.169034	0.653149
s33	tatuadores	male tattoo artists	68.750	1.803	0.006999	0.006999
s34	inventores	male inventors	68.750	1.803	0.921496	0.085538
s35	ministros	male ministers	68.281	1.568	2.383978	1.383855
s36	millionarios	male millionaires	68.281	1.936	0.718804	0.349215
s37	políticos	male politicians	68.125	1.371	2.326735	2.752374
s38	exploradores	male explorers	68.095	1.808	0.878337	0.120053
s39	diputados	male representatives	67.656	1.692	1.750337	1.002679
s40	neurocirujanos	male neurosurgeons	67.656	1.663	0.089001	0.006999

(continued on next page)

Sequence no.	Spanish role noun	English translation	Male-biased rating score	Standard error	LogFreq masculine	LogFreq feminine
s47	caseros	male homemade products makers	47.419	2.444	0.660919	0.577638
s48	veterinarios	male veterinarians	47.031	1.632	0.574267	0.619920
s49	niñitos	male grandchildren	46.984	1.688	0.126428	0.239126
s50	vecinos	male neighbors	45.968	1.407	1.521806	1.256515
s51	humanitarios	male humanitarians	45.625	1.589	0.858274	1.265642
s52	camareros	male waiters	45.469	1.558	0.722295	0.525354
s53	vendedores	male vendors	45.313	2.004	0.982735	0.387279
s54	biólogos	male biologists	45.000	1.606	0.519421	0.121122
s55	famosos	male celebrities	44.844	1.510	1.702011	1.488339
s56	ginecólogos	male gynecologist	44.375	2.596	0.229251	0.044235
s57	compradores	male customers	42.188	1.816	0.826853	0.172420
s58	pasteleros	male bakers	40.161	2.089	0.160883	0.107016
s59	farmacéuticos	male pharmacists	39.524	2.164	0.696681	0.741650
s60	profesores	male professors	38.125	1.650	1.906708	1.047162
s61	mecanógrafos	male typists	37.343	2.162	0.028666	0.044235
s62	psicopedagogos	male educational	36.250	2.107	0.011144	0.000297
s63	bibliotecarios	male librarians	33.125	1.709	0.509494	0.181805
s64	coreógrafos	male choreographers	32.656	1.970	0.470703	0.212277
s65	cosmetólogos	male cosmeticians	32.064	2.764	0.001409	0.001409
s66	modistos	male tailors	31.406	2.052	0.100347	0.324531
s67	adivinos	male fortune tellers	30.469	1.921	0.471657	0.886293
s68	cajeros	male cashiers	29.375	2.281	0.362889	0.177138
s69	bailarines	male dancers	26.250	1.703	0.693263	0.831661
s70	peluqueros	male hairdressers	24.531	1.589	0.299778	0.123252
s71	tejedores	male weavers	23.438	1.578	0.422273	0.094711
s72	cuidadores	male caretakers	22.188	1.659	0.638801	0.116830
s73	azafatos	male flight attendants	21.875	1.589	0.002813	0.314388
s74	enfermeros	male nurses	21.094	1.357	0.370772	0.811010
s75	secretarios	male secretaries	21.094	1.741	2.057915	1.267170
s76	animadores	male cheerleaders	20.938	2.013	0.578384	0.275028
s77	maquilladores	male make-up artists	20.469	1.908	0.024685	0.029985
s78	limpiadores	male cleaners	18.594	1.424	0.092436	0.077349
s79	prostitutos	male prostitutes	16.094	1.329	0.045507	0.654088
s80	niñeros	male baby sitters	15.938	1.489	0.019319	0.456141