Sally the Congressperson: The Role of Gender Ideology in Gender-Neutral Neologism Processing

B. T. Papineau

June 2021

Word Count: 3,052

LINGUIST 245B: Methods in Psycholinguistics

1 Introduction

It is well-established in the sociolinguistic and psycholinguistic literature that interlocutors rapidly integrate social information about both their world and their interlocutor in the course of language processing. For example, Casasanto (2008) found that comprehenders integrated social information about talkers (in this instance, race) in their processing of variable t/d deletion, and reaction times were significantly faster when t/d deletion was paired with a Black face vesus a white one. Similarly, Van Berkum et al. (2008) found ERP evidence which indicates that speech produced by a speaker for whom the utterance is atypical (for example, a child saying "I quit drinking alcohol last week") invokes more processing difficulty than speech produced by a typical speaker. These examples serve to show that interlocutors integrate both speaker information (such as race and age) and real-world information (the association between age and drinking, for example) in real-time sentence processing. However, not all findings have been congruent in this manner; von der Malsburg et al. (2020) found that the female pronoun 'she' saw no processing advantage even by participants who believed that their respective female candidates were expected to win in the 2016 US Presidential Election and the 2017 UK Parliamentary Election. They also found that the female pronoun showed a slight advantage in the British case when coreferring with the future Prime Minister, but not in the American context.

The integration of extra-linguistic information in the course of language comprehension is formalized in Levy (2008)'s Surprisal Theory, captured in the notion of extrasentential context, which informs the relative surprisal of word in a particular sentential context. Despite this formalization, however, relatively little work has looked at the role of individual differences in ideologies and linguistic exposure as a factor in establishing the relative weight and power of extrasentential contexts in surprisal theory. That being said, some recent research has begun to examine the role of individual differences in lexical decision and retrieval tasks (Davies et al., 2017; Monaghan et al., 2017; Yap et al., 2008).

This project sets out to synthesize these separate lines of inquiry, by applying a surprisal-theory based approach to investigate how individuals' ideologies related to gender and gender roles may inform the relative surprisal values of a particular gendered or gender-neutral term in an inherently gendered context. In order to do this, we conducted a web-based self-paced reading experiment, wherein participants saw sentences such as *David is a congressman/person* or *Sally is a congresswoman/person*. According to surprisal theory, we might expect that the gender of the character in the vignette (male or female) will influence the processing time on the critical item, in this case the professional title, if the individual holds gender-based beliefs about who should be fulfilling the particular social, economic, or political role. This should in turn be reflected in a longer reading time, where reading time is representative of processing power required for a particular syntactic item (Monsalve et al., 2012).

We might logically hypothesize that individuals with more open-minded approaches to gender will be more likely to allow women to perform historically male-dominated jobs, and vice versa, than their more gender-traditional counterparts. We might also assume that there is still a bias in society which centers men in the workplace. Based on these assumptions, we make three explicit hypotheses about what we expect to find in our experiment:

1. An individual with more open-minded gender ideologies should be faster to process gender-neutral terms combined with gendered names, as a result of weaker prior beliefs about who should be performing that role. This is contrasted with gender-traditional individuals, who we expect to see a penalty in such cases. This prediction is bolstered by the assumption that

individuals likely seek out media which reflects their values, and therefore may be more (in the case of open-minded individuals) or less (in the case of gender-traditional individuals) exposed to the gender neutral forms. Such exposure should translate to individual frequency effects, where it is well-established that higher frequency items are processed more quickly than lower-frequency items, possibly as a result of their being relatively less surprising (Brysbaert et al., 2016; Monsell et al., 1989).

2. Given both real-world probabilities and heterosexism in the workplace (the U.S. Department of Labour (2020) estimates that 53% of the American workforce is male), we expect that critical items proceeded by male-coded names will be processed more easily (and therefore more quickly) than the same items in female-coded contexts.

3. Finally, given the neologistic nature of many of the compounds in the semantic fields of person and professional titles, resulting in lower frequencies for these terms compared to their gendered counterparts, we expect that gender-neutral contexts will be processed more slowly than their gender-congruent counterparts.

After conducting the self-paced reading study (explicated in Section 2), we find some support for Hypothesis 2, but none for either Hypothesis 1 or Hypothesis 3; these results are reported in Section 2.2. Finally, reasons for this and limitations of the current study are discussed in Section 3.

2 Experiment 1

2.1 Methodology

All experimental materials, including the experiment, stimuli, and analysis scripts, can be found in the author's public GitHub repository at: https://github.com/BranPap/gender_processing under the '1_Reading_Neutrality' subfolders. We also provide here a link to our preregistration for this experiment: https://osf.io/jwx7k.

2.1.1 Participants

200 pariticipants were recruited through the online participant recruitment platform Prolific. All participants were native speakers of English, were born and resided in the United States, and had not participated in the norming study. The mean completion time of the study was 5.38 minutes. All participants, regardless of final inclusion in data analysis (see below) were paid $1.75 USD for their participation in the experiment, for an average payout rate of $20.73/hr.

Participants who scored less than 85% accuracy on attention check questions were excluded from analysis[1]. This left us with a total of 186 participants (mean age: 32.69 years; F: 98; M: 81; Other: 4; N/A: 3) whose data was included in the analysis.

2.1.2 Materials

Participants saw twenty pairs of sentences, which were presented one word at a time. There were 20 critical lexemes, each with (up to) 3 gendered forms: a male form (e.g. *congressman*), a female form

[1] 13 of these 14 exclusions were because of incompatibilities between the experiment and the FireFox internet browser. Subsequent experiments (not reported in this paper) instructed participants not to use this browser.

(e.g. *congresswoman*), and a gender neutral variant (e.g. *congressperson*). Paired with the binary genders of male and female, this provides six potential configurations, but the gender-incongruent mapping (e.g. *David* + *congresswoman*) was excluded, because it was feared that it would bring too much attention to the critical issue of gender.

As such, there were four possible sentence configurations for each of the critical items; sentences could be either **congruent** in their character-noun gender mapping or **neutral** in their character-noun gender mapping. Examples are provided in Table 1. Each sentence was also paired with a random activity and location (US State)- these information then served the basis of the attention check questions. The states and activities were consistent across all participants for an individual lexeme (for example, all participants always saw the congressperson/woman/man as being from Virginia and enjoying cycling). The names of the characters assigned to each profession were randomized within and between participants, in an attempt to control for any biases that might arise from individual names. Names were selected from the Top 100 male and female names from the author's birth year (1998), and any names which were gender-ambiguous (e.g. *Ryan*) or whose spelling might be misinterpreted (e.g. *Daniel* being read as *Danielle*) were excluded.

Full example pairs from the stimuli are provided below:

Condition	**Example Stimulus**
Male-Congruent	*David is a congressman from Virginia. He likes cycling.*
Male-Neutral	*David is a congressperson from Virginia. He likes cycling.*
Female-Congruent	*Sally is a congresswoman from Virginia. She likes cycling.*
Female-Neutral	*Sally is a congressperson from Virginia. She likes cycling.*

Table 1: Example stimuli from experiment

All critical items, along with 19 others (for a total of 39 items) were normed prior to experimentation. 100 participants (recruited the same way, and paid at the same rate as in Section 2.1.1.) were provided with sentences such as *Someone is a congressperson*, and asked to indicate on a Likert scale how likely they thought it was that the unnamed character was a woman or man. However, because the female forms caused significant colinearity within our mixed-effects model, these norming scores were excluded from the statistical analyses. We do, however, employ them in our analyses of individual lexical items.

Both the main experiment and the norming study were written in HTML, JavaScript, and CSS, and were hosted on the author's GitHub Pages website, where it was completed by participants.

2.1.3 Procedure

The experiment consisted of four phases. In the first phase, participants were told that they would be reading sentences one word at a time, and that they would have to answer questions about each of the sentence pairs that they read. They were then provided with an example sentence (*Brittany is a performer from South Korea. She likes cycling*), and asked to answer a question about the sentence (*Is Brittany from Lithuania?*) which they had to proceed through by pressing the spacebar. All participants saw the same example sentence and were asked the same question, which they responded to by clicking on buttons labeled *Yes* and *No*. If participants responded incorrectly to the question, they were not able to proceed until they had corrected their choice; feedback on their selection was provided.

The second phase of the experiment consisted of the main experimental trial. Participants were presented with 20 sentences, in a randomized order, constructed in the manner outlined in Section 2.1.2. All sentences were displayed as a series of dashes on the screen, and pressing the spacebar converted the next set of dashes into a word, and then back to dashes when the spacebar was pressed again and the next word was revealed. At the end of each sentence pair, participants were randomly asked about the preferred activity or the US state of origin of the character in that sentence pair. The time between spacebar presses was recorded, as were the name in question, the lexeme, the specific orthographic form, the condition, and the trial number.

After completing the 20 main trials, participants responded to the 13-question Social Roles Questionnaire put forth in Baber and Tucker (2006). This happened over two pages: in the first, participants were asked 5 questions about their beliefs regarding gender as something that is inherent and important to be maintained as social categories- these questions are scored on a sliding scale of 0 to 100, where a 0 indicates a more conservative approach to gender, and a 100 indicates a more open-minded approach to gender. These questions correspond to Baber and Tucker (2006)'s Gender Transcendence subscore, and were inversely coded in analysis. On the second page, participants were asked 8 questions pertaining to the characteristics of and social roles performed by the binary male and female genders, which correspond to Baber and Tucker (2006)'s Gender Linked subscale. These were likewise scored from 0 to 100, but in this case a 0 represented a more open-minded approach to gender roles, whereas a 100 represented a more strict belief system regarding those same roles. Participants had to respond to all 13 questions in order to continue, and all scales were presented in a gender-neutral color (green) and with no starting marker on the scale, so as to not bias participants one way or the other.

Finally, participants were given the opportunity to provide demographic information. This included gender, age, and political alignment, as well as any additional comments or feedback on the payment or enjoyability of the experiment. All parts of this last section were optional, though the vast majority of participants responded to most or all of the questions, with the exception of the free-response feedback.

2.2 Results

All subsequent analyses were carried out in the open-source statistical package R (R Core Team, 2021), and trial gender (male vs. female) and trial congruency (congruent vs. incongruent) were centered. Additionally, gender ideology (the means score of all 13 questions in the Social Roles Questionnaire) was scaled and centred. All results are reported in residual reading times, after running a simple linear regression to control for the effect of orthographic length on reading times.

To assess the role of gender ideology on the processing of both gendered and gender-neutral nominals, we conducted a mixed-effects linear regression using the R package lme4 (Bates et al., 2015) wherein we predicted residual reading time as a function of trial gender, trial congruency, gender ideology, and the interactions between these three predictors. Random by-item and by-participant intercepts were included, as well. Through this analysis, we found that there was a significant effect of the gender of the trial (β = -4.477e-02, SE = 1.048e-02, t=-4.271, $p < .001$) such that nominals paired with male names were processed more quickly than nominals paired with female names. This appears to be driven primarily by an interaction between trial gender and trial congruency such that gender-congruent forms paired with male names were read more quickly than any other gender-congruency condition (β = 4.567e-02, SE = 2.097e-02, t=0.0295, $p < .05$); this is evident in Figure 1, where the average reading time for the congruent male condition is significantly

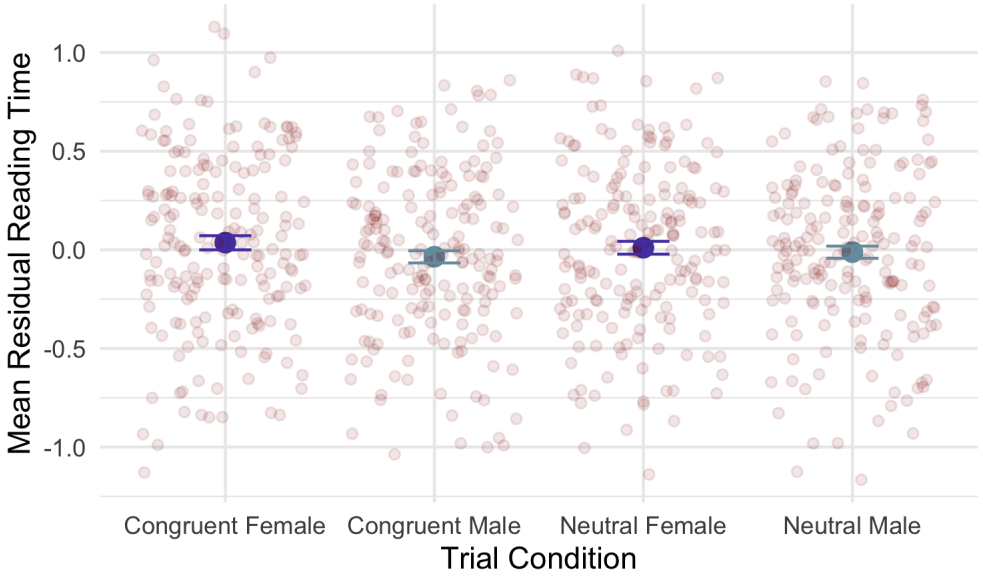

Figure 1: Mean residual reading time by trial condition across 186 participants. Red dots represent individual participants' means in the relevant condition.

quicker than the other three conditions.

Moreover, there appears to be an effect of gender ideology that is trending towards significance (β = -5.129e-02, SE = 2.948e-02, t=-4.271, $p < .1$), whereby more gender-conservative individuals read the critical items more quickly than gender-open individuals. This trend can be observed in Figure 2. We discuss possible reasons for this in Section 3, as well as confounding factors that may have given rise to this correlation.

Finally, we do find a great amount of variation within individual lexemes, as shown in Figure 3. For example, we find that in the neutral congruency trials, "flight attendant" is read significantly faster when paired with a female name than with a male name. On the other hand the opposite is true in the "congressperson" case; "congressperson" is read significantly faster when paired with a male name compared to a female name.

3 Discussion

3.1 General

Our results seem to indicate that there is a significant facilitation effect when reading professional and personal titles paired with male names. Crucially, however, this does not appear to be mediated at all by an individual's gender ideology as measured by the Social Roles Questionnaire (Baber & Tucker, 2006); rather, there is only a broad finding that gender-conservative individuals read critical items more quickly than gender-open individuals. Therefore, on the basis of the findings herein, we do not find support for Hypothesis 1 as stated in Section 1, which states that individuals with more

Figure 2: Mean residual reading time by gender ideology; a score of 0 indicates an open-minded approach to gender, while a score of 100 would indicate a conservative approach to gender.

open-minded approaches to gender should be quicker at processing gender-neutral forms, either because their own priors about the gender of a profession's agent are less rigid or as a result of media exposure to neologistic forms. Rather, it seems that these professional and personal titles are simply less surprising when paired with a male name than with a female name (though see below for apparent counterexamples); as a result, we can take this as support for Hypothesis 2, which predicted that male titles will be processed more quickly than female titles.

We also do not find support for Hypothesis 3, which predicted that neologistic neutral forms (such as "congressperson") should be read more slowly than their gendered counterparts. This seems to indicate that they are either not sufficiently infrequent or surprising to require additional processing power.

Despite these findings, there are several remaining concerns which cast doubt on the results. We attempt to address these in turn here.

First and foremost, there is a significant skew in our sample towards the open-minded end of the gender ideology scale employed; the mean gender ideology score was 20.66 out of 100, and the maximum extended only to 60.24. As a result, it is difficult for us to draw any significant conclusions about the role of gender ideology on the processing of these terms. As a result of this, a second experiment, identical to the first, was conducted with 100 self-identified Republicans on Prolific, who we hypothesised would have higher scores on the gender ideology scale, indicating more conservative social ideologies. The results of this experiment, however, are not included in this analysis, and will be included in later permutations of this work. There is also a secondary confound with the gender ideology score, which is that participants always filled this out *after* the main trials. As such, they were exposed to both men and women performing a variety of social, political, and economic roles, including in roles that exist in counter to real-world probabilities (e.g.

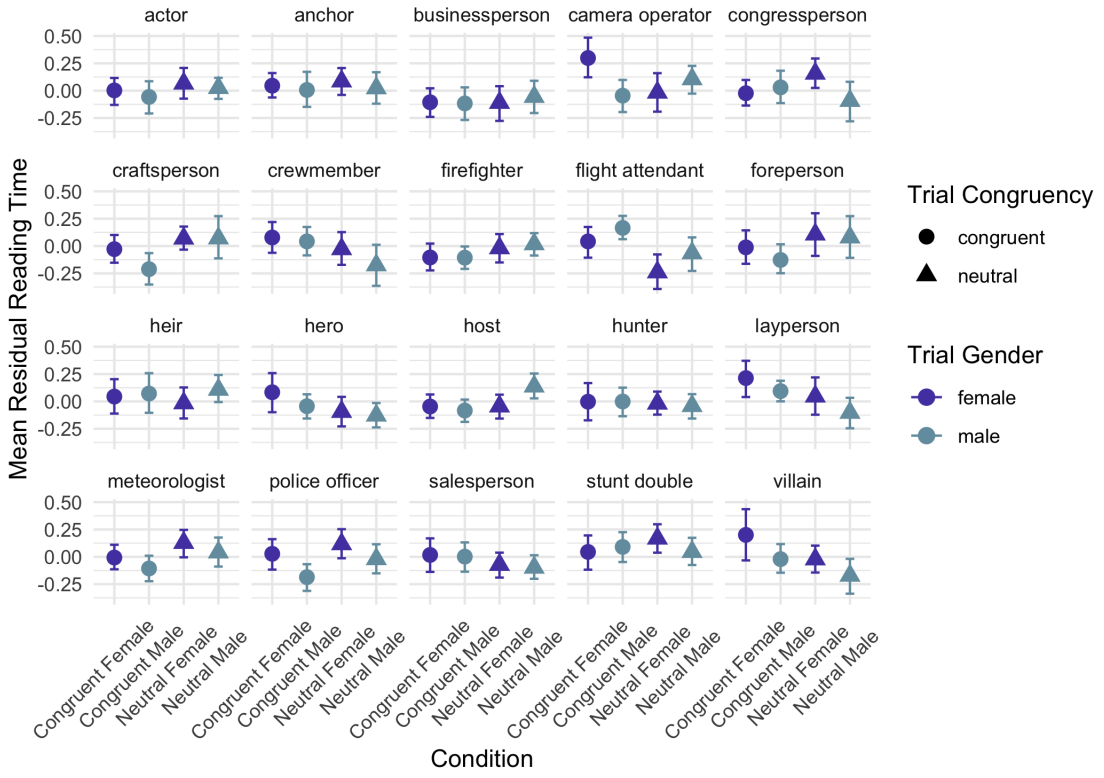

Figure 3: Mean residual reading time for individual lexical items, by trial congruency and trial gender.

female congresspeople, while in reality the US Congress is less than 27% female). This may have primed participants to respond as having a more socially-open gender ideology.

Moreover, there is not yet an implementation in our mixed-effects model of lexical frequency or surprisal ratings (as estimated by language models such as gpt2 or bert). Beacuse of this, it is entirely that some of the results we are seeing are a result of particular items being extremely infrequent (such as 'camerawoman') and/or surprising. This is the next step of the project, and we are currently working on modelling these surprisal ratings for inclusion in future work.

3.2 Lexeme-Specific

Despite these limitations, there are interesting item-specific findings which bear mentioning, especially in the gender-neutral contexts. For example, returning to Figure 3, we see that in both the gender-neutral *and* the gender-congruent conditions, female flight attendants (column 4, row 2) are read more quickly than male flight attendants. This reflects real-world probabilities, wherein more than 75% of flight attendants are women (Data USA, 2019), and also socially-entrenched stereotypes. Conversely, in the congruent context at least, male congresspeople are processed more

quickly than female congresspeople. Again, this reflects real-world probabilities and stereotypes. Interestingly, in the cases which are not true neologisms, but rather semantic broadening of masculine terms to include female referents in the sets they denote (*actor, heir, hero, host, hunter, villain*), there is relatively little difference in processing between the genders in congruent contexts; only villain and host show significant differences. This may indicate that these terms have reached a more constant state of gender-neutrality than their neologistic counterparts, but again more information about relative surprisal is needed in order to tease this apart.

4 Conclusion

In this paper, we report on the findings of a self-paced reading study, which seem to indicate that there is not a significant effect of individuals' gender ideology on the processing of gendered and gender-neutral personal and professional titles. However, in line with surprisal theory (Levy, 2008), we do find that individual items' reading times reflect real-world probabilities and heterosexist social ideologies, which can be understood as extrasentential contexts. Despite these findings, however, more work with more conservative individuals, as well as with modelled surprisal values, is needed in order to complete the picture sketched herein.

References

Baber, K. M., & Tucker, C. J. (2006). The social roles questionnaire: A new approach to measuring attitudes toward gender. *Sex Roles*, *54*(7-8), 459–467.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the dutch lexicon project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441.

Casasanto, L. S. (2008). Does social information influence sentence processing? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*(30).

Current population survey 2020. (2020). US Bureau of Labor Statistics. https://www.dol.gov/agencies/wb/data/latest-annual-data/working-women#Labor-Force-by-Sex-Race-and-Hispanic-Ethnicity

Davies, R. A., Arnell, R., Birchenough, J. M., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1298.

Flight attendants. (2019). Data USA. https://datausa.io/profile/soc/flight-attendants#demographics

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Monaghan, P., Chang, Y.-N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, *93*, 1–21.

Monsalve, I. F., Frank, S. L., & Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 398–408.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, *118*(1), 43.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of cognitive neuroscience*, *20*(4), 580–591.

von der Malsburg, T., Poppels, T., & Levy, R. P. (2020). Implicit gender bias in linguistic descriptions for expected events: The cases of the 2016 united states and 2017 united kingdom elections. *Psychological science*, *31*(2), 115–128.

Yap, M. J., Balota, D. A., Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by rt distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 495.