Why Produce Garden-Path Sentences? Evidence from Corpus Data across Registers

Rei Emura (rei.emura.r4@dc.tohoku.ac.jp)[1,2], **Stefan L. Frank (stefan.frank@ru.nl)**[2],
Saku Sugawara (saku@nii.ac.jp)[3,4] **& Masatoshi Koizumi (koizumi@tohoku.ac.jp)**[1]
[1]Graduate School of Arts and Letters, Tohoku University
[2]Centre for Language Studies, Radboud University
[3]National Institute of Informatics, Japan
[4]The University of Tokyo

Abstract

Garden-path sentences induce processing difficulty by requiring readers to revise an initial syntactic analysis (e.g., *The policeman saw the lights were off*). While garden-path comprehension has been widely studied, far less is known about why such sentences are produced in natural language. We thus address this gap by analyzing large-scale written-text corpora in English and Japanese across five registers. English Noun/Sentence garden-path constructions, arising from optional complementizer omission, were frequent in blogs and newspaper, suggesting that their emphasis on communicative efficiency promotes their use. Rather, Japanese Main Clause/Relative Clause garden-path constructions, which arise from grammatical constraints, were most frequent in legal texts, reflecting the register's tendency to favor complexity within a clause. Legal texts also exhibit the longest ambiguous regions across constructions, which can increase reanalysis cost. These findings suggest that grammatical properties and register-specific writing styles jointly shape the production of garden-path sentences in real-world language.

Keywords: garden-path; language production; corpus study; register

Introduction

Successful communication requires avoiding the production of sentences that are difficult to comprehend. Garden-path (GP) sentences are known to incur larger processing costs and misinterpretations because the processor frequently constructs an incorrect analysis mid-sentence, requiring a reanalysis process (Frazier, 1983). For example, in (1), readers first assume a simple meaning in which the policeman saw some lights. Encountering *were* later in the sentence forces them to revise this interpretation and realize that the policeman saw that the lights were off.

(1) *The policeman saw the lights were off.*

Even though the comprehension process has been widely studied (for reviews, see Frances, 2024), how to apply this insight to our everyday communication has not been well discussed. Several studies have failed to find evidence that speakers produce or omit *that*, as in (1), in order to reduce temporal ambiguity for the listener (V. S. Ferreira & Dell, 2000; Jaeger, 2010; Roland et al., 2006). This suggests that speakers do not naturally take the listener's incremental processing into account. Therefore, it is important to examine what leads people to produce GP sentences, as avoiding GPs would be useful for successful and smooth communication.

This study examines how frequently and under what situations people produce GP sentences and when more difficult GPs are produced in real language use. Then, through exploratory data analysis, we aim to discuss why people produce GP sentences and more difficult ones.

Garden-Path Types

A wide range of GP constructions has been studied across languages. The present study focuses on GP types that differ in how they are produced, distinguishing between **optionally abbreviated GP constructions** and **structurally natural GP constructions**. As summarized in Table 1, we examine one English GP type classified in the former group: NP/S (Noun Phrase/Sentence). In the latter group, we examine the Japanese MC/RC (Main Clause/Relative Clause) GP type.

In English, NP/S GPs arise through the optional omission of grammatical markers. In NP/S constructions, omitting the complementizer *that* leads the parser to initially analyze a postverbal noun phrase as a direct object, which is later reanalyzed as the subject of an embedded clause (Garnsey et al., 1997; Sturt, 2007, among others).

On the other hand, Japanese MC/RC GPs arise from grammatical constraints rather than optional omission. Japanese lacks overt relative clause markers, and relative clauses obligatorily precede the head noun. Consequently, a verb phrase following a matrix subject can initially be interpreted as a matrix predicate but must later be reanalyzed as a relative clause modifying a subsequent noun (Mazuka, 1991; Mazuka & Itoh, 2013, among others).

This distinction between optional and structurally natural GP constructions is expected to have important consequences for their distribution across registers. In the two subsections below, we derive these register-based predictions.

Disambiguation Difficulty

Disambiguation difficulty is affected by the length of ambiguous regions. It has been demonstrated that a longer ambiguous region leads to more difficulty in the reanalysis process in English (F. Ferreira & Henderson, 1991; Tabor & Hutchins, 2004; Van Dyke & Lewis, 2003; Warner & Glass, 1987) and Japanese (Arai & Nakamura, 2016; Emura et al., 2025; Nakamura & Arai, 2016).

Second, we also consider surprisal at the disambiguation region as an index of the processing difficulty of GPs. Surprisal follows from the probability of a given word given the

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Table 1: Target constructions.

Language	Type	Garden-Path Construction	Preferred Construction
English	NP/S	*The policeman saw **the lights** <u>were</u> off.*	*The policeman saw the lights.*
Japanese	MC/RC	*Pianisuto-ga [$_{RC}$**piano-o hiita**] <u>bareriina</u>-o mitsuketa.* piannist-SUB [$_{RC}$piano-OBJ played] ballerina-OBJ found "The pianist found the ballerina who played the piano."	*Pianisto-ga piano-o hiita.* pianist-SUB piano-OBJ played "The pianist played the piano."

Note. Bold text and underlined text represent ambiguous regions and disambiguation regions, respectively. We use the term "preferred construction" to refer to the construction consistent with the initial parsing of GP sentences. NP/S = Noun Phrase/Sentence; MC/RC = Main Clause/Relative Clause; SUB = subject; OBJ = object; RC = relative clause.

context, defined as Surprisal $= -\log P(\text{word}|\text{context})$ (Hale, 2001; Levy, 2008), which means that less predictable words have larger surprisal. Surprisal influences the processing difficulty at the disambiguation region, in combination with the length of ambiguous regions (Emura et al., 2025), although Arehalli et al. (2022) and Van Schijndel and Linzen (2021) demonstrated that Surprisal does not solely explain the difficulty of reanalysis.

Syntactic Difference by Register

Even though it has not been demonstrated in which specific situations GP sentences are used, there are general differences in syntactic structure across registers (Conrad, 2023, for review). This study focuses on five written registers: Legal, News, Literature, Magazine, and Blog, which systematically differ in their degree of formality. In particular, Legal and News texts are highly formal, whereas Blogs are relatively informal (Larsson, 2019).

One relevant syntactic phenomenon is the omission of complementizers in that-clauses: the complementizer *that* is most frequently omitted in conversational language, followed by fiction and news reportage, whereas academic prose exhibits the lowest omission rates (Biber & Conrad, 2005). This distribution leads to predictions about NP/S ambiguities. Specifically, among the registers examined in the present study, NP/S GP sentences are expected to occur more frequently in Blog texts, which are closer to conversational language, and less frequently in informational registers such as Legal documents and News articles.

At the same time, informational written registers are characterized by a tendency toward complex structures in noun-phrase constituents and clauses (Biber & Gray, 2010; Biber et al., 2011). From this perspective, Japanese MC/RC structures are predicted to occur more frequently in Legal and News texts, where dense modification and explicit clause-level information are common.

Finally, given the properties of Legal text, which exhibits longer and more complex structures within a clause, we predict that ambiguous regions will be longer in this register, regardless of language or GP type.

The Current Study

Taken together, the current study investigates how frequently GP sentences are produced in natural written language and how that depends on disambiguation difficulty . Using large-scale text corpora, we quantified the frequency and structural properties of these constructions across five written registers: Legal, News, Literature, Magazine, and Blog.

Based on previous findings on register variation and syntactic omission, we formulated three primary hypotheses:

(i) Abbreviated GP constructions (i.e., English NP/S) are expected to occur less frequently in highly formal registers and more frequently in less formal registers.

(ii) In contrast, structurally natural GP constructions (i.e., Japanese MC/RC) are expected to show the opposite pattern.

(iii) GP sentences in Legal texts are expected to exhibit longer ambiguous regions, regardless of GP construction.

Methods

Corpora

For English, we used the Corpus of Contemporary American English (COCA, 25+ million words, between 1990–2019; Davies, 2008–), focusing on four genres: news, magazines, fiction,[1] and blogs. To represent legal texts, we additionally included the Constitution of the United States (from 1787) and a portion of the Federal Laws (between 2005–2024). For Japanese, we used the Balanced Corpus of Contemporary Written Japanese (BCCWJ, 104.3 million words, between 1976-2006; Maekawa et al., 2014), selecting texts from five registers: legal, newspaper, magazines, book, and blogs.

Target Constructions

As shown in Table 1, we examined the distribution of GP candidates. We use the term "candidate" because we extract sentences whose constructions may cause garden-path effects.We defined GP templates and two baselines: (i) a *preferred baseline* with a similar lexical configuration but matching the preferred initial analysis, and (ii) general sentences used as a reference distribution.

[1] In this study, we collectively refer to the register labeled *fiction* in COCA and *book* in the BCCWJ as *literature*.

NP/S Ambiguity in English. English NP/S GP candidates were defined as *Matrix-Subject + Matrix-Verb + (no complementizer) + Subordinate-Subject + Subordinate-Verb*. The extraction criteria were: (i) the matrix clause contains a subject and verb but no noun-phrase object; (ii) the matrix verb is restricted to 189 lemmas that take sentential complements (e.g., *see*, *think*, *find*); (iii) the sentence contains a subordinate clause whose subordinate subject is adjacent to the matrix verb, with no subordinating conjunction, punctuation, or complementizer *that*; and (iv) the subordinate subject does not change its form depending on whether it is a subject or object (e.g., *he*, *she*, *I*, and *who*). The ambiguous region spanned from the subordinate subject up to but not including the subordinate verb; the disambiguation region was the subordinate verb.

The preferred baseline was defined as follows: (i) the sentence contains a matrix subject, a matrix object, and a matrix verb; (ii) the matrix verb is restricted to the same 189 lemmas that take sentential complements; and (iii) the matrix object does not change its form depending on whether it is a subject or object.

MC/RC Ambiguity in Japanese. Japanese MC/RC GP candidates were defined as *Matrix-Subject + Relative-Object + Relative-Verb + Matrix-Object*. The criteria were: (i) the sentence contains a matrix subject, object, and verb; (ii) the matrix object serves as the head noun modified by a relative clause; (iii) the relative clause consists of a relative object and verb. The ambiguous region was the relative clause; the disambiguation region was its head noun.

As a preferred baseline, we extracted sentences that (i) contained a matrix subject, object, and verb, and (ii) in which matrix object was not modified by a relative clause.

Analysis

Preprocessing. All texts were processed using spaCy, an open-source library for Natural Language Processing (Honnibal et al., 2020). First, we segmented the corpus into sentences based on the following criteria: (i) the sentence contained at least one period; (ii) the sentence length was at least 10 tokens; (iii) the sentence did not include more than three consecutive punctuation marks; and (iv) duplicate sentences were removed. For preprocessing, we used a sentencizer pipeline[2] for English, and *ja_core_news_lg*[3] for Japanese.

Main Analysis. We parsed all the sentences, and counted the number of target sentences and the number of tokens in the ambiguous regions.[4] We used transformer-based spaCy pipelines: *en_core_web_trf*[5] and *ja_core_news_trf*[6] for English and Japanese, respectively. For each type, we created 10 test sentences and verified that all of them were classified correctly.

We additionally calculated Surprisal of the disambiguation regions. We used GPT-2-small, one of the pre-trained neural language models (for English, Radford et al., 2019; for Japanese, Sawada et al., 2024).

We then statistically tested register effects on the proportions of GPs using a binomial generalized linear model, and on ambiguous region length and Surprisal using linear models. We assessed the statistical significance using an ANOVA on the fitted models, and then conducted post-hoc pairwise comparisons between registers using Tukey-adjusted estimated marginal means. We used R (R Core Team, 2025), and the *emmeans* package (Lenth, 2025) for pairwise comparisons.

Results

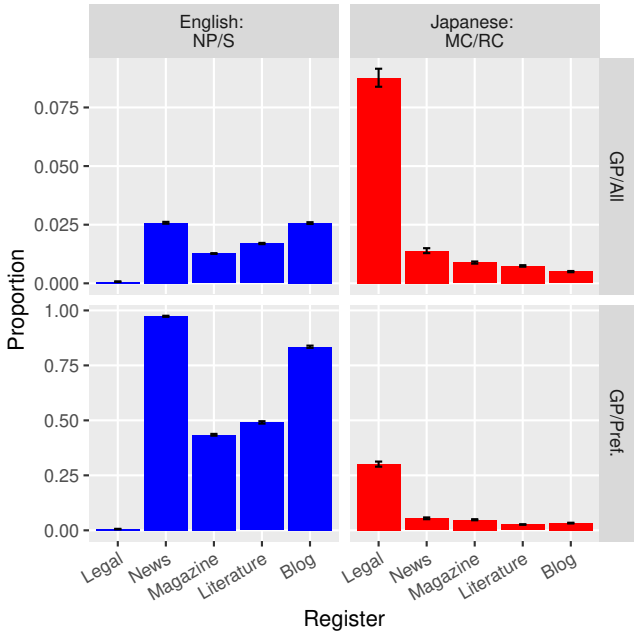

Figure 1: Proportion of garden-path candidates across registers. The graph represents the ratio of GP candidates in each register, using two measures: GP/All (the proportion of garden-path candidates among all sentences) and GP/Unamb. (the proportion of garden-path candidates relative to the preferred baseline construction).

Table 2 shows examples of extracted GP candidates. Overall, the frequency of GP candidates varied substantially across registers and constructions. For each construction, we report two frequency measures: the proportion of GP candidates among all sentences and relative to the number of sentences with preferred construction (see Table 3 and Figure 1). In addition, Figure 2 reveals the number of tokens in the ambiguous region and surprisal at the disambiguating token.

[2]https://spacy.io/api/sentencizer

[3]https://spacy.io/models/ja#ja_core_news_lg

[4]In SpaCy, a token can be a word, punctuation symbol, whitespace, etc. (https://spacy.io/api/token)

[5]https://spacy.io/models/en#en_core_web_trf

[6]https://spacy.io/models/ja#ja_core_news_trf

Table 2: Examples of extracted GP candidates.

GP Type	Register	Extracted Sentence
English: NP/S	News	*A toxicology report revealed Ward was under the influence of marijuana at a level high enough to cloud judgment.*
English: NP/S	Blog	*Second, we all know the whole TIVA relationship is completely dysfunctional.*
Japanese: MC/RC	Legal	*... tachiiri-kensa-no kengen-wa [$_{RC}$hanzai-sosa-no tame-ni* ... on-site-inspection-GEN authority-SUB [$_{RC}$ criminal-investigation-GEN for *mitome-rare-ta] mono-to kaishi-te-wa-nara-nai.* be-granted-for] something-as shall-not-construe "... the authority for the on-site inspections shall not be construed as something which is granted for criminal investigation."

GP = garden-path; SUB = subject; OBJ = object; RC = relative clause; GEN = genitive case.

Table 3: Frequency of target constructions, length of ambiguous regions, and Surprisal at disambiguation regions.

Register	All sentences	Pref.	GP	GP/All	GP/Pref.	Length	Surprisal
English: NP/S							
Legal	158,805	19,538	110	0.000693	0.005630	9.636 (0.913)	3.519 (0.226)
News	618,952	16,410	15,975	0.025810	0.973492	3.946 (0.031)	5.036 (0.026)
Magazine	1,922,646	56,212	24,422	0.012702	0.434462	3.847 (0.028)	5.075 (0.023)
Literature	929,302	32,137	15,772	0.016972	0.490774	2.926 (0.023)	5.441 (0.029)
Blog	685,136	21,078	17,599	0.025687	0.834946	3.557 (0.028)	4.807 (0.027)
Japanese: MC/RC							
Legal	20,766	6,046	1819	0.087595	0.300860	20.352 (0.388)	2.677 (0.050)
News	47,374	12,087	658	0.013889	0.054439	14.769 (0.326)	4.950 (0.128)
Magazine	156,869	28,820	1387	0.008842	0.048126	13.508 (0.220)	5.478 (0.092)
Literature	267,603	74,777	1981	0.007403	0.026492	11.589 (0.157)	6.233 (0.083)
Blog	286,603	43,544	1430	0.004989	0.032840	12.573 (0.200)	5.417 (0.097)

Note. Length and surprisal values are reported as means, with standard errors in parentheses. GP: Garden-path construction; Pref.: Preferred construction; GP/All: the proportion of garden-path candidates among all sentences; GP/Pref.: the proportion of garden-path candidates relative to the preferred baseline construction.

English: NP/S

For the English NP/S ambiguity, GPs were most frequent in *News* and *Blog* registers, whereas *Legal* texts show the lowest GP frequency (Figure 1). Specifically, the proportion of GPs within all sentences is *News* = *Blog* ⋙ *Literature* ⋙ *Magazine* ⋙ *Legal*.[7] Similarly, the proportion of GPs relative to the preferred construction is *News* ⋙ *Blog* ⋙ *Literature* ⋙ *Magazine* ⋙ Legal.

Furthermore, the number of tokens in the ambiguous region was the largest in the legal text, specifically, *Legal* ⋙ *News* = *Magazine* ⋙ *Blog* ⋙ *Literature* (Figure 2). The order of Surprisal is *Literature* ⋙ *Magazine* = *News* ⋙ *Blog* ⋙ *Legal*. It reveals that Surprisal shows an almost inverse of ambiguous length: shorter ambiguous regions tended to show higher Surprisal.

[7]We use the following symbols to indicate the significance level of differences in proportion: >: $p < 0.05$; ≫: $p < 0.01$; ⋙: $p < 0.001$; =: non significant.

Japanese: MC/RC

For Japanese MC/RC structures, GPs are most frequent in Legal texts in contrast to English NP/S construction (Figure 1). Specifically, the order of GP/All is *Legal* ⋙ *News* ≫ *Magazine* ⋙ *Literature* ⋙ *Blog*, and the order of GP/Preferred is *Legal* ⋙ *News* ⋙ *Magazine* ⋙ *Blog* ⋙ *Literature*.

The ambiguous regions are the longest in Legal, same as English NP/S (Figure 2). In particular, the order of ambiguous length is *Legal* ⋙ *News* = *Magazine* = *Blog* = *Literature* (*News* ⋙ *Blog* and *News* ⋙ *Literature*). Besides, the distribution of Surprisal shows a nearly reverse pattern of ambiguous length, similar to English NP/S: *Literature* ⋙ *Magazine* = *Blog* > *News* ⋙ *Legal*.

Discussion

The present study aims to investigate why people produce GP sentences, and examines how frequently GP sentences are produced in natural written language and how their distribution varies across registers in English and Japanese. From the

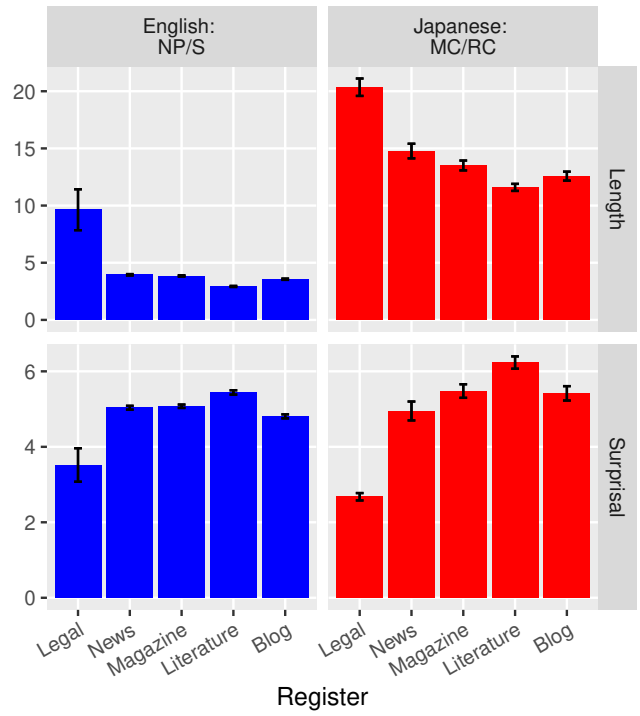

Figure 2: Mean number of tokens in the ambiguous regions and mean Surprisal at the disambiguation region. Error bars represent 95% confidence intervals.

corpus analyses, we discuss the findings corresponding to our hypotheses: with regard to the hypotheses (i) and (ii), the writing motivations of each register lead to the (un)production of GPs. With respect to hypothesis (iii), properties of Legal texts systematically extend ambiguous regions, thereby increasing disambiguation difficulty. Details are provided below.

Production of GP Structures

From the properties of registers, we discuss why people produce GP sentences, and argue that the reasons differ depending on the GP construction. With respect to the frequency of GP sentences, the results provide converging evidence that GP production is shaped by an interaction between the grammatical properties of a language and register-specific production pressures. More specifically, we argue that production constraints such as limited space and rapid writing promote the use of "abbreviated" GP constructions in English (i.e., NP/S ambiguity), whereas stylistic preferences for detailed explanation within a single clause promote the use of "structurally natural" GP constructions in Japanese (i.e., MC/RC ambiguity).

English: NP/S. Within English, the results partially support hypothesis (i) that GP constructions arising from the optional omission of overt grammatical markers occur less frequently in formal registers and more frequently in informal ones. Specifically, the data show that NP/S GP sentences were frequent in Blogs and least frequent in the Legal register. These results can be attributed to their relatively conversational writing style: the complementizer *that* is more likely to be omitted in conversational language (Biber & Conrad, 2005), creating NP/S ambiguity. Legal writing occupies the formal end of the register spectrum, leading to the avoidance of complementizer omission.

In addition, the News register unexpectedly shows the highest proportion of GP constructions. This suggests that production constraints other than conversationality are also at play. In particular, newspapers are often subject to space limitations, which may encourage more compact sentence structures and increase complementizer omission.

Taken together, these factors suggest that limited space and rapid writing can promote the production of NP/S GP sentences. From the perspective of communicative efficiency, although shorter expressions may increase the risk of temporary ambiguity, writers may still adopt them when efficiency is prioritized over explicitness (Piantadosi et al., 2012). Our results suggest that the generation of NP/S GPs is driven by the prioritization of efficiency over explicitness.

Japanese: MC/RC. For Japanese, the data support hypothesis (ii) that structurally natural GP constructions occur more frequently in informational registers. Specifically, Japanese MC/RC GPs were overwhelmingly most frequent in Legal texts, followed by the News register, and least frequent in Blogs. This pattern can be understood in light of stylistic differences across registers. Biber and Gray (2010) and Biber et al. (2011) suggested that informational documents, such as academic prose, tend to favor dense modification within a single clause, especially in noun phrases. Such a writing style naturally promotes the use of relative clauses, which in Japanese can give rise to GP structures due to the absence of overt relative clause markers. As a result, efforts to provide detailed and explicit descriptions within a single clause may inadvertently increase the likelihood of MC/RC GPs.

Production of GP Sentences with Disambiguation Difficulty

With respect to disambiguation difficulty, consistent with hypothesis (iii), Legal texts consistently exhibited the longest ambiguous regions, regardless of GP type or language. This parallel pattern across English and Japanese suggests that a property of informational texts, which permit complex and highly detailed explanations within a single clause (Biber & Gray, 2010; Biber et al., 2011), can lengthen ambiguous regions and increase the potential processing cost. Thus, this property of Legal text not only increases the frequency of Japanese MC/RC GPs, but also increases reanalysis cost across GP types.

Surprisal is an index of general processing cost. At the level of registers, we observed an inverse relationship between surprisal and the length of the ambiguous region, meaning that registers with longer length have lower surprisal. We additionally examined the direct relationship between ambigu-

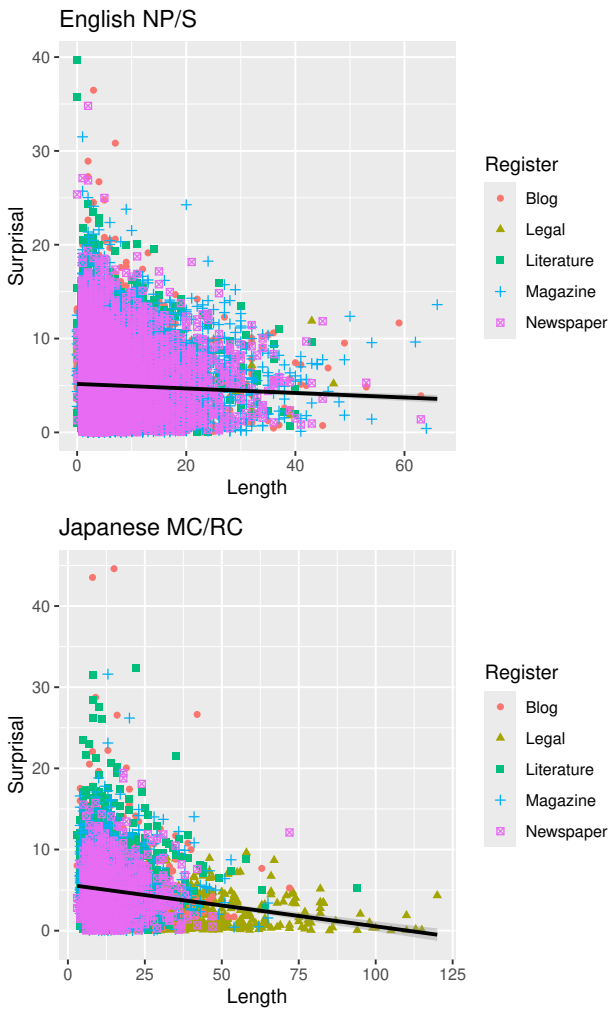

Figure 3: The negative relationship between ambiguous region length and surprisal in English NP/S and Japanese MC/RC garden-path sentences. The lines show linear trends; shaded regions indicate standard errors.

ous region length (i.e., context length) and surprisal. As shown in Figure 3, longer ambiguous regions were generally associated with lower surprisal. Specifically, ambiguous region length had a significant negative effect on surprisal in the linear mixed-effects model, with register included as a random intercept, both in English NP/S ($\beta = -0.02, SE = 0.00, t = -5.90, p < 0.001$) and in Japanese MC/RC ($\beta = -0.02, SE = 0.00, t = -4.38, p < 0.001$).

This pattern likely reflects a general property of large language models, whereby longer preceding context leads to lower surprisal at subsequent words (Futrell et al., 2020; Hahn et al., 2022; McCurdy & Hahn, 2024). Crucially, however, lower surprisal in registers with longer ambiguous regions does not imply reduced disambiguation difficulty because the effects of ambiguous region length on the reanalysis process can persist independently of surprisal (Emura et al., 2025). Our data did not show that surprisal increased or decreased in a register-specific manner after accounting for the length of the ambiguous region.

Contributions of this Study

This study contributes to the garden-path literature by shifting the focus from comprehension to production. Whereas previous research has primarily examined how garden-path sentences are processed, the present study provides corpus-based evidence for when and why such sentences are produced in natural written language. Our results show that the production of garden-path sentences is systematically shaped by an interaction between grammatical constraints and register-specific writing pressures.

We also refine our understanding of the uniqueness of Legal texts. Legal writing is known to involve structures associated with high processing demands, such as longer syntactic dependencies (Martínez et al., 2022, 2023). This study expands this property to GP sentences: their syntactic complexity promotes longer ambiguous regions when such sentences occur, which are expected to increase reanalysis costs. At the same time, GP constructions that depend on optional omission, such as English NP/S, are comparatively rare. Together, these findings show that Legal texts selectively combine grammatical complexity with stylistic constraints, rather than uniformly increasing all forms of processing difficulty.

Limitations and Future Direction

The occurrence of NP/S GPs (i.e., *that*-omission) is known to depend on the probability that a given matrix verb licenses or is followed by *that* Jaeger, 2010. In the present study, we did not control for or model verb-specific biases associated with matrix verbs. As a result, variation in the likelihood of *that*-omission across verbs may have influenced the observed patterns. Future work should incorporate verb-level properties, such as complementizer bias.

Another limitation concerns data quality. Because GP candidates were extracted using automatic parsing by spaCy, some sentences were incorrectly classified due to parsing errors, even though we confirmed that the test data was correctly classified. Moreover, whether extracted GP candidates actually induce a garden-path effect depends not only on syntactic configuration but also on lexical-semantic factors (Christianson et al., 2001, 2006; Nakamura & Arai, 2016). Future research should combine corpus-based extraction with behavioral validation or more fine-grained semantic controls.

Conclusion

This study demonstrates that the production of GP sentences is affected by an interaction between grammatical systems and register-specific production goals. By situating classic GP constructions within naturalistic corpora, we bridge experimental psycholinguistics and real-world language use, highlighting how both the frequency and the severity of garden-path effects are shaped by communicative context.

References

Arai, M., & Nakamura, C. (2016). It's harder to break a relationship when you commit long. *Plos one*, *11*(6), e0156482. https://doi.org/10.1371/journal.pone.0156482

Arehalli, S., Dillon, B., & Linzen, T. (2022). Syntactic surprisal from neural models predicts, but underestimates, human processing difficulty from syntactic ambiguities. In A. Fokkens & V. Srikumar (Eds.), *Proceedings of the 26th conference on computational natural language learning (conll)* (pp. 301–313). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.conll-1.20

Biber, D., & Conrad, S. (2005). Register variation: A corpus approach. In *The handbook of discourse analysis* (pp. 175–196). John Wiley Sons, Ltd. https://doi.org/10.1002/9780470753460.ch10

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in l2 writing development? *TESOL Quarterly*, *45*(1), 5–35. https://doi.org/10.5054/tq.2011.244483

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults'" good-enough" interpretations of garden-path sentences. *Discourse processes*, *42*(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6

Conrad, S. (2023). *Corpus Linguistics and Linguistic Theory*, *19*(1), 7–21. https://doi.org/10.1515/cllt-2022-0032

Davies, M. (2008–). *The corpus of contemporary american english (coca)*. https://www.english-corpora.org/coca/

Emura, R., Kawachi, Y., Sugawara, S., & Koizumi, M. (2025). The lingering effect as memory persistence has distinct predictors from the garden-path effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *Advance online publication*. https://doi.org/10.1037/xlm0001538

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745. https://doi.org/10.1016/0749-596X(91)90034-H

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340. https://doi.org/https://doi.org/10.1006/cogp.1999.0730

Frances, C. (2024). Good enough processing: What have we learned in the 20years since ferreira et al. (2002)? *Frontiers in Psychology*, *Volume 15 - 2024*. https://doi.org/10.3389/fpsyg.2024.1323700

Frazier, L. (1983). Processing sentence structures. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 215–236). Academic Press. https://doi.org/10.1016/B978-0-12-583680-7.X5001-2

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, *44*(3), e12814. https://doi.org/10.1111/cogs.12814

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58–93. https://doi.org/10.1006/jmla.1997.2512

Hahn, M., Futrell, R., Levy, R., & Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, *119*(43), e2122602119. https://doi.org/10.1073/pnas.2122602119

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the Second Meeting of the North American Chapter of the Asssociation for Computational Linguistics*, 159–166. https://aclanthology.org/N01-1021/

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *Spacy: Industrial-strength natural language processing in python*. https://doi.org/10.5281/zenodo.1212303

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/https://doi.org/10.1016/j.cogpsych.2010.02.002

Larsson, T. (2019). Grammatical stance marking across registers. *Register Studies*, *1*(2), 243–268. https://doi.org/10.1075/rs.18009.lar

Lenth, R. V. (2025). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.11.2-8]. https://doi.org/10.32614/CRAN.package.emmeans

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., & Den, Y. (2014). Balanced corpus of contemporary written japanese. *Language Resources and Evaluation*, *48*, 345–371. https://doi.org/10.1007/s10579-013-9261-0

Martínez, E., Mollica, F., & Gibson, E. (2022). Poor writing, not specialized concepts, drives processing difficulty in legal language. *Cognition*, *224*, 105070. https://doi.org/10.1016/j.cognition.2022.105070

Martínez, E., Mollica, F., & Gibson, E. (2023). Even lawyers do not like legalese. *Proceedings of the National Academy of Sciences*, *120*(23), e2302672120. https://doi.org/10.1073/pnas.2302672120

Mazuka, R. (1991). Processing of empty categories in japanese. *Journal of psycholinguistic research*, *20*(3), 215–232. https://doi.org/10.1007/BF01067216

Mazuka, R., & Itoh, K. (2013). Can japanese speakers be led down the garden path? In *Japanese sentence processing* (pp. 295–329). Psychology Press.

McCurdy, K., & Hahn, M. (2024). Lossy context surprisal predicts task-dependent patterns in relative clause processing. In L. Barak & M. Alikhani (Eds.), *Proceedings of the 28th conference on computational natural language learning* (pp. 36–45). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.conll-1.4

Nakamura, C., & Arai, M. (2016). Persistence of initial misanalysis with no referential ambiguity. *Cognitive Science*, *40*(4), 909–940. https://doi.org/10.1111/cogs.12266

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. https://doi.org/10.1016/j.cognition.2011.10.004

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*.

Roland, D., Elman, J. L., & Ferreira, V. S. (2006). Why is that? structural prediction and ambiguity resolution in a very large corpus of english sentences. *Cognition*, *98*(3), 245–272. https://doi.org/https://doi.org/10.1016/j.cognition.2004.11.008

Sawada, K., Zhao, T., Shing, M., Mitsui, K., Kaga, A., Hono, Y., Wakatsuki, T., & Mitsuda, K. (2024). Release of pretrained models for the Japanese language. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 13898–13905. https://aclanthology.org/2024.lrec-main.1213

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*(2), 477–488. https://doi.org/10.1016/j.cognition.2006.10.009

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 431–450. https://doi.org/10.1037/0278-7393.30.2.431

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316. https://doi.org/10.1016/S0749-596X(03)00081-0

Van Schijndel, M., & Linzen, T. (2021). Single-stage prediction models do not explain the magnitude of syntactic disambiguation difficulty. *Cognitive science*, *45*(6), e12988. https://doi.org/10.1111/cogs.12988

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language*, *26*(6), 714–738. https://doi.org/10.1016/0749-596X(87)90111-2