

The distribution and variation of non-coordinated pronoun case forms in English

Tyler Lemon*

Abstract. This paper investigates case preferences for English pronouns in non-coordinated environments subject to variation: in isolation and after *not*, with following numeral, PP, and NP modifiers, in *than* and *as* comparatives, and as the foci of *it*-clefts. It uses an acceptability judgment experiment (Mechanical Turk) and corpus study (COCA) to investigate these preferences and presents an Optimality Theory-style analysis based on their results that models structural case assignment (case from an external head), default case, and prescriptivism as competing constraints. This model captures case preferences and their relative strengths and provides one possible explanation for the variation attested in these environments.

Keywords. pronoun; case; variation; non-coordinated; English; Optimality Theory; acceptability judgment experiment; Mechanical Turk; corpus study; COCA

1. Introduction. Any native speaker of English who has thought significantly about his or her language would agree that a pronoun's case is often not simply a matter of its function as subject or object. Perhaps the most attention has been given to coordinated pronouns (1a), which vary their case in a way that lone pronouns cannot (1b). Other authors' examples use their judgments.

- (1) Coordination (Grano 2006:1)
 - a. {She/Her} and Sandy went to the store yesterday.
 - b. {She/*Her} went to the store yesterday.

There are also some non-coordinated environments that allow variation between nominative (*I, he, she, we, they*) and accusative (*me, him, her, us, them*) forms. Examples include pronouns with following NP modifiers (ignoring the argument-modifier distinction) (2), the foci of *it*-clefts (3), *than* comparatives (4), and at least for *I*, following *not* in an isolated utterance (5).

- (2) NP modifier (Schütze 2001:215)
{We/Us} linguists are a crazy bunch.
- (3) *It*-cleft
It was {?she/her} that won the contest.
- (4) *Than* comparative
Sarah is taller than {?I/me}.
- (5) After *not*
Who took the cookies? - Not {?I/me}.

These examples reveal that there are a number of environments that allow both nominative and accusative case forms. This variation is an interesting puzzle for linguistic theory. Why do these environments not require a specific case? This question is certainly not fully resolved, but current theories (Quinn 2002, Schütze 2001, Parrott 2006, Grano 2006, and Sobin 1997 among

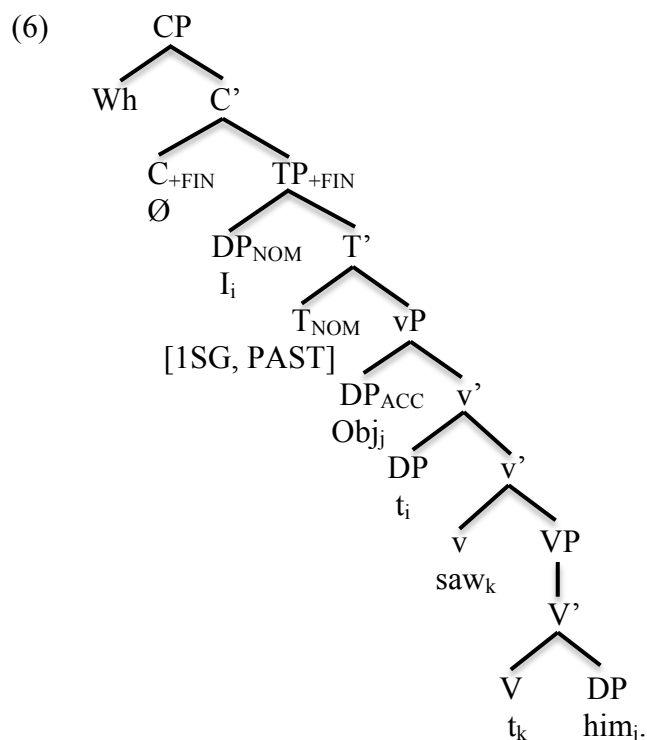
* I would like to thank Lelia Glass, Arnold Zwicky, and especially Tom Wasow for all of their help on this project. Author: Tyler Lemon, University of California, Berkeley (tylerlemon@berkeley.edu). This is a much shorter version of my undergraduate honors thesis at Stanford University. Please email me for a copy of the long version.

others) attribute these patterns to a combination of prescriptivism, default case, and which environments unambiguous case assignment can penetrate.

In this paper I provide experimental data on case in non-coordinated environments in which speakers exhibit variation, and I also produce an analysis of this data within the context of contemporary case theory that captures the relative strengths of the factors noted above and their interactions in the different environments. This paper is organized into six sections, including the introduction. Section 2 provides an overview of the different types of case, focusing on those applicable to English. Section 3 describes the methodology and results of the acceptability judgment experiment performed for this paper. Section 4 presents the methodology and results of an accompanying corpus study. Section 5 discusses the results, including broader trends, and provides a theoretical analysis. Section 6 concludes the paper with a summary of its significance and speculation about unsolved problems with suggestions for future research.

2. Types of case. This section provides an overview of the different kinds of case and notes the types applicable to English. To start, linguists commonly divide case into at least two types: structural and non-structural case. To this default case and prescriptive case may be added.

2.1. STRUCTURAL CASE. Structural case is associated with particular syntactic positions and is not assumed to be controlled by lexical heads (like V and P). More general aspects of the grammar (such as functional heads in many frameworks) control and license it. As an illustrative example, Chomsky (1995:369) assumes the structure in (6) for the transitive sentence *I saw him*.



In this structure, the subject *I* (the external argument) is generated as a specifier of *v* (Spec-*v*P) and moves to Spec-TP to check nominative case. The object *him* (the internal argument) is generated as the complement of *saw*. *Him* does not move overtly to Spec-*v*P to check accusative, but its features do (Hornstein, Nunes & Grohmann 2005:120). *Saw* is generated in V and moves overtly to *v* due to features on *v* that attract a lexical V to it. Crucially, case assignment in this sentence is handled by functional heads, T (working with C) for the subject and *v* for the object.

2.2. NON-STRUCTURAL CASE. Non-structural case can be split into inherent case and lexical case. Inherent case is associated with particular θ -roles/positions, examples being ergative and dative case when they fill their “predictable” functions as an external argument and recipient/goal respectively (Woolford 2006:111-112). The latter is illustrated by the Icelandic ditransitive in (7), where the indirect object recipient *henni* receives the typical dative case. Lexical case is unpredictable, idiosyncratic case assigned by lexical heads, an example being the unpredictable genitive assigned to the direct object theme *Þess* (these are usually accusative in the language).

- (7) Icelandic (Zaenen, Maling & Thráinsson 1985:470)
 Þú hefur óskað henni Þess
 you.SG.NOM have wished her.DAT this.GEN
 ‘You have wished her this.’

I would argue that English lacks inherent and lexical case. Prescriptive influences aside, case on whole-DP pronouns in English is completely predictable by structural position. English no longer distinguishes a dative, and there are no lexical items that assign “idiosyncratic” case. To show the absence of non-structural case in English, one common test to distinguish the two broad types is case preservation under A-movement (passives, raising). Non-structural case is preserved, but structural case is not (Woolford 2006:117-118). In English, case on subjects and objects of all types is not preserved in passivization (8). By contrast, Icelandic preserves inherent datives (9).

- (8) a. I gave her the cookie.
 b. She was given the cookie by me.
 c. They saw her.
 d. She was seen by them.
 e. We spoke to him.
 f. He was spoken to by us.
- (9) Icelandic (Zaenen, Maling & Thráinsson 1985:471)
 henni var óskað Þess
 her.DAT was wished this.GEN
 ‘She was wished this.’

2.3. DEFAULT CASE. Another type of case not covered above but which is certainly present in English is default case. Default case is in a sense a last resort mechanism that assigns case when no other part of the grammar can do so, permitting the licensing of all non-PRO DPs (PRO can only bear null case, Martin 2001:146) in isolated answers (10) and other default case environments without a licensing head. In English, the default is accusative (Quinn 2002:70), and as the judgment experiment will show, there is typically a strong preference for accusative in environments where structural case is not assumed to apply.

- (10) a. Who saw Michael? - {*I/Me}.
 b. Who did Michael see? - {*I/Me}.

The environments in which default case applies are language-specific. In German, for example, isolated-answer DPs match the case of the corresponding Wh-word in a preceding question (11). At least according to Schütze (2001:208), default case is probably available to all languages, but its implementation varies. Understanding default case in a language requires a detailed comprehension of the environments where other mechanisms predominate and where they do not.

- (11) German (Frazier 2007:9)
- a. Wer hat Hans geschlagen? - {Ich/*mich}.
 who.NOM has Hans hit - {I/*me}
 ‘Who hit Hans? - {I/*me}.’
- b. Wen hat Hans geschlagen? - {*Ich/mich}.
 who.ACC has Hans hit - {*I/me}.
 ‘Who did Hans hit? - {*I/me}.’

2.4. PRESCRIPTIVISM. The last factor that influences case on pronouns in English is prescriptivism, a term designating the forms of language that grammarians would consider “correct” or “proper”. According to Sobin (1997), the nominative variants of (12a)-(12c) are examples of “linguistically deviant” prestige constructions; the accusative variants are non-prestige.

- (12) Prestige constructions (Sobin 1997:318)
- a. Mary and {I/me} left early.
 b. It was {I/me}.
 c. Mary is richer than {I/me}.

The three environments above are ones in which nominative *I* is prescriptively correct, but accusative *me* is either nearly as possible (12a) or generally preferred outside of formal contexts (12b and 12c). It often sounds much more natural to use an informal variant, but prescriptivist teachings are common from a young age in formal education, and this can leave a lasting mark on how speakers view particular variants. However, because prestige variants are often artificially imposed on the grammar that native speakers of English acquire naturally as children, we usually fail to fully internalize these rules, and there are at least three pieces of evidence for this.

The first is that prescriptive rules must be explicitly taught. Many rules, such as nominative for whole-DP pronoun finite subjects (1b), are nearly categorical in English, but even after years of education in prestige usage, speakers often make “mistakes” with modified pronouns (2) and those in the environments in (12), and they must be corrected repeatedly.

The second is that prescriptive rules are often incompletely learned, and their application is overextended (Sobin 1997:331). Prescriptivists complain of phrases like (13), arguing that nominatives should not be used within verbal objects. (13) is an instance of hypercorrection, the overextension of a prestige form into an environment where it was not intended to apply.

- (13) Hypercorrection (Huddleston & Pullum 2002:463)
 %They’ve invited the Smiths and we to lunch.

The third is that many prescriptive rules are specific to certain lexical items and go against trends in English grammar. As an example, *than* can take both TP and DP complements, and accusative is preferred on the DPs. This makes the case that *than* is a C/P hybrid (Hankamer 1973) like *before* and *after*, but only *than* has prescriptively stipulated nominative objects (14) (Zwicky 2012).

- (14) a. You came {before/after} I came.
 b. You came {before/after} {I/me}.

This section demonstrates that there are (at least) three distinct mechanisms at play in the determination of pronoun case forms in English: structural case, default case, and prescriptivism. The first two are part of the grammar proper of English, while the third is artificially imposed and extralinguistic, but a relevant factor nevertheless. This competition is not always obvious, but this paper has noted a number of environments where case assignment becomes more ambigu-

ous, and the interaction of these mechanisms becomes clearer. With this knowledge of the relevant mechanisms in place, sections 3 and 4 investigate many of the environments discussed above and several others through an acceptability judgment experiment and corpus study.

3. Acceptability judgment experiment. The following section details the methodology, results, and data analysis of the acceptability judgment experiment conducted for this paper.

3.1. **METHODOLOGY.** The acceptability judgment experiment was conducted on Mechanical Turk and Qualtrics. Mechanical Turk was used to advertise the survey to workers, but the actual survey was on Qualtrics; a link to the Qualtrics survey was given to workers who accepted the HIT (Human Intelligence Task) on Mechanical Turk. 200 native speakers of American English, 196 of whom actually finished the survey, rated 34 sentences for their acceptability on a 7-point Likert scale. Participants were instructed to rate the sentences (either a standalone or an answer to a question) based on their “naturalness”, not what they may have been taught as correct in school. The 14 stimuli each had multiple pronoun variants, which were distributed randomly to a roughly equal number of participants. Pronoun case was never mentioned, and the 20 unrelated fillers were inserted in an effort to direct attention away from what was really being examined. The sentences were presented one at a time in a randomized order. Participants were paid \$1.50 for approximately 10 minutes of work, regardless of how much of the survey they completed.

3.2. **STIMULI AND RESULTS.** The following examples (15-28) and corresponding Tables (1-14) present the stimuli and the average ratings given for each pronoun in these environments.

(15) Who took the cookies? - {I/he/we/me/him/us}.

Case	1SG	3SG	1PL
NOM	3.03	2.70	2.09
ACC	6.32	5.61	4.52

Table 1: In isolation – Subject

(16) Who did Robert give the book to? - {I/he/we/me/him/us}.

Case	1SG	3SG	1PL
NOM	2.53	1.79	2.06
ACC	6.67	5.41	5.58

Table 2: In isolation – Object

(17) Who took the cookies? - Not {I/he/we/me/him/us}.

Case	1SG	3SG	1PL
NOM	5.52	2.73	2.21
ACC	6.75	5.94	6.18

Table 3: After *not* – Subject

(18) Who did Robert give the book to? - Not {I/he/we/me/him/us}.

Case	1SG	3SG	1PL
NOM	4.70	2.29	2.03
ACC	6.42	5.18	5.67

Table 4: After *not* – Object

(19) {You/We/Us} two enjoyed the movie.

Case	2PL	1PL
NOM	5.78	3.29
ACC		2.78

Table 5: Numeral modifier – Subject

(20) Mark really relies on {you/we/us} two.

Case	2PL	1PL
NOM	6.31	2.21
ACC		5.26

Table 6: Numeral modifier – Object

(21) {You/We/Us} from California have to deal with a lot of traffic.

Case	2PL	1PL
NOM	2.91	3.86
ACC		2.68

Table 7: PP modifier – Subject

(22) Traffic is a fact of life for {you/we/us} from California.

Case	2PL	1PL
NOM	3.15	2.63
ACC		4.78

Table 8: PP modifier – Object

(23) {You/We/Us} Californians get the best weather.

Case	2PL	1PL
NOM	6.05	5.98
ACC		4.65

Table 9: NP modifier – Subject

(24) Life can be very fast-paced for {you/we/us} Californians.

Case	2PL	1PL
NOM	6.09	3.65
ACC		6.11

Table 10: NP modifier – Object

(25) Harry is taller than {I/he/we/me/him/us}.

Case	1SG	3SG	1PL
NOM	4.56	3.39	2.18
ACC	6.36	5.94	6.44

Table 11: *Than* comparative

(26) Harry is as tall as {I/he/we/me/him/us}.

Case	1SG	3SG	1PL
NOM	4.52	3.24	3.03
ACC	6.18	5.25	5.55

Table 12: *As* comparative

(27) Do you think that John won the contest?

- No, it was {I/you/she/we/they/me/her/us/them} that won.

Case	2	1SG	3SG	1PL	3PL ¹
NOM	5.19	4.50	3.19	2.82	3.23
ACC		5.09	4.22	4.50	3.13

Table 13: *It*-cleft – Subject

(28) Did Tyler see Samantha in the garden?

- No, it was {I/you/she/we/they/me/her/us/them} that Tyler saw.

Case	2	1SG	3SG	1PL	3PL
NOM	4.50	3.23	2.00	2.86	2.95
ACC		4.95	2.95	5.25	3.73

Table 14: *It*-cleft – Object

3.3. DATA ANALYSIS. To determine the overall case preferences in each syntactic environment, the scores given to each stimulus by a given participant were first converted to z-scores¹ using the fillers for standardization. Each filler had only one variant, and they were designed to range from completely acceptable to completely unacceptable to measure how much of the scale each participant used and if their ratings skewed towards one end or the other. Each participant's average filler score and filler score standard deviation were calculated, and z-scores for their stimuli were measured as standard deviations from their average filler score. This provided a form of standardization that controls at least partially for potential biases in scale use.

Due to space considerations, Table 15, which summarizes the case preferences in each environment and the average difference in ratings between nominatives and accusatives on the Likert scale, is provided below. An explanation for how these results were derived follows.

Environment	Preferred case / Average difference	
	Subject	Object
In isolation	ACC / 2.88	ACC / 3.75
After <i>not</i>	ACC / 2.80	ACC / 2.75
Numeral modifier	NOM / 0.50	ACC / 3.05
PP modifier	NOM / 1.18	ACC / 2.15
NP modifier	NOM / 1.33	ACC / 2.46
<i>Than</i> comparative	-	ACC / 2.87
<i>As</i> comparative	-	ACC / 2.06
<i>It</i> -cleft	ACC / 0.80	ACC / 1.46

Table 15: Summary of the results of the acceptability judgment experiment

¹ After conversion to z-scores (see section 3.3), the average rating given to *them* became higher than for *they*.

After deriving the z-scores, those for the nominative and accusative variants were compared (excluding case-neutral *you* when tested) using a Welch’s t-test, which estimates whether two sets of data with unequal sample sizes and unequal variances have significantly different means. Because of the way Qualtrics administered the survey, some variants of a given stimulus received a few more ratings than others, and some received no ratings at one end of the scale or the other. Ultimately, all case preferences were highly statistically significant ($p < 0.001$) except the preference for nominative of subject numeral modifiers ($p \approx 0.22$).

4. Corpus study. This section details the methodology, results, and analysis of the corpus study.

4.1. **METHODOLOGY.** The corpus study was conducted on Mark Davies’ online Corpus of Contemporary American English (COCA), a corpus that consists of approximately 534 million words of written and spoken American English dating from 1990 to 2015. The purpose of this study was to attempt to quantify the influence of prescriptivism on English pronoun case. As such, it focuses on two relevant environments with enough tokens for statistical analysis: after *than* and after *not*. The pronouns were specified to have following punctuation (. or ,) to ensure that they did not begin TPs. Crucially, COCA can conduct searches by genre (spoken, fiction, magazine, newspaper, and academic). Formality in English typically entails a greater use of nominative pronouns (Huddleston & Pullum 2002:460), so one would expect a higher ratio of nominatives in more formal genres like academic writing than in less formal ones like spoken.

4.2. **RESULTS.** Tables 16-27 provide the number of tokens of each pronoun in a given environment and genre. *3SGM/F* indicates 3rd singular masculine/feminine, and *Pro* the pronoun.

Case	1SG	3SGM/F	1PL
NOM	93	65/28	37
ACC	1703	388/255	300

Table 16: *Not Pro ./*, – All of COCA

Case	1SG	3SGM/F	1PL
NOM	9	8/3	7
ACC	347	92/52	79

Table 17: *Not Pro ./*, – Spoken

Case	1SG	3SGM/F	1PL
NOM	5	8/1	3
ACC	191	49/11	58

Table 18: *Not Pro ./*, – Newspaper

Case	1SG	3SGM/F	1PL
NOM	6	6/3	5
ACC	201	37/19	46

Table 19: *Not Pro ./*, – Magazine

Case	1SG	3SGM/F	1PL
NOM	48	35/16	20
ACC	913	206/169	97

Table 20: *Not Pro ./*, – Fiction

Case	1SG	3SGM/F	1PL
NOM	25	8/5	2
ACC	51	4/4	20

Table 21: *Not Pro ./*, – Academic

Case	1SG	3SGM/F	1PL
NOM	400	317/273	62
ACC	1502	291/184	291

Table 22: *Than Pro ./*, – All of COCA

Case	1SG	3SGM/F	1PL
NOM	56	19/8	15
ACC	265	56/47	49

Table 23: *Than Pro ./*, – Spoken

Case	1SG	3SGM/F	1PL
NOM	31	21/12	5
ACC	182	48/19	75

Table 24: *Than Pro ./*, – Newspaper

Case	1SG	3SGM/F	1PL
NOM	53	29/18	4
ACC	229	35/15	48

Table 25: *Than Pro ./*, – Magazine

Case	1SG	3SGM/F	1PL
NOM	231	217/217	27
ACC	788	148/100	101

Table 26: *Than Pro ./*, – Fiction

Case	1SG	3SGM/F	1PL
NOM	29	31/18	11
ACC	38	4/3	18

Table 27: *Than Pro ./*, – Academic

4.3. DATA ANALYSIS. The first step in analyzing the data was to calculate the percentages of nominative and accusative pronouns relative to all tokens for both environments in all the genres. The percentages of nominatives are provided in Tables 28 and 29.

Condition	Spoken	Newspaper	Magazine	Fiction	Academic	All
1SG	2.53	2.55	2.90	4.99	32.89	5.18
3SGM	8.00	14.04	13.95	14.52	66.67	14.34
3SGF	5.45	8.33	13.64	8.65	55.56	9.89
1PL	8.14	4.92	9.80	17.09	9.09	10.98
All	4.52	5.21	6.19	7.91	33.61	7.77

Table 28: *Not Pro ./*, – % NOM

Condition	Spoken	Newspaper	Magazine	Fiction	Academic	All
1SG	17.44	14.55	18.79	22.67	43.28	21.03
3SGM	25.33	30.43	45.31	59.45	88.57	52.14
3SGF	14.55	38.71	54.55	68.45	85.71	59.74
1PL	23.44	6.25	7.69	21.09	37.93	17.56
All	19.03	17.57	24.13	37.83	58.55	31.69

Table 29: *Than Pro ./, – % NOM*

In both environments, academic writing had by far the highest percentage of nominatives, followed by fiction, and then magazine writing. After *not*, spoken examples have the smallest percentage of nominatives, and newspaper writing the second least; after *than* the reverse is true.

To verify that genre is a relevant factor in these differences, Pearson’s chi-squared tests were performed comparing the number of nominative and accusative tokens in each genre (not the percentages). This test determines the likelihood that differences observed in categorical data (NOM vs. ACC) according to a categorical independent variable (genre) have arisen by chance. When comparing all genres, genre was found to be a highly significant factor in determining the ratio of NOM to ACC ($p < 1.00 \cdot 10^{-25}$). When comparing each genre to the ones adjacent to it in terms of this ratio (academic vs. fiction, fiction vs. magazine, etc.), all differences were statistically significant except for that between magazine and fiction writing after *not* ($p \approx 0.07$).

Apart from the one exception noted above, the statistical analysis confirms the qualitative finding that more formal genres (especially academic writing) have a significantly higher ratio of nominatives in English in environments subject to prescriptive influence. All of this indicates that prescriptivism exerts a strong influence on English pronoun case when it is most active. Therefore, a full account of the facts of English pronoun case must incorporate it.

5. Discussion and analysis. This section summarizes case preferences, discusses some important trends that emerge from the results of the acceptability judgment experiment and corpus study, and provides an OT-style theoretical analysis to synthesize all these results.

5.1. SUMMARY OF PREFERENCES. The results in Table 15 show that overall, participants in the acceptability judgment experiment preferred accusative pronouns in isolation and after *not* (assumed default case environments) by a significant margin, at least 2.5 points on the Likert scale in all four cases. The case preferences for the pronouns with numeral, PP, and NP modifiers matched the structural case that would be assigned unambiguously to an unmodified pronoun; subjects were judged more acceptable on average with nominative case, and objects exhibited a preference for accusative case. However, the preference for nominative of subject pronouns with numeral modifiers was small (≈ 0.50 points) and thus not statistically significant. Accusative pronouns were favored in *than* and *as* comparatives (≥ 2 points), as would be expected if *than* and *as* are prepositions assigning structural accusative. Finally, accusatives were preferred as the foci of *it*-clefts (another assumed default case environment), though the margins were often small (< 1 point), especially when the pronoun was linked to a subject in the lower clause (Table 13).

The corpus study broadly corroborated the results of the acceptability judgment experiment in that the number of accusative tokens after *not* and *than* outnumbered nominative tokens by a large margin (92.23% and 68.31% ACC tokens respectively). It also demonstrated the significant effect of genre and thus the importance of prescriptivism in determining pronoun case in English.

5.2. GENERAL TRENDS. Examining the results of the judgment experiment especially, two notable trends emerge: the fact that, comparing scores by case, accusative pronouns were always rela-

tively more acceptable in subject position than nominative pronouns were in object position and the fact that *I* was routinely rated more acceptable than other nominative pronouns where tested.

5.2.1 THE RELATIVE ACCEPTABILITY OF ACCUSATIVE SUBJECTS AND NOMINATIVE OBJECTS. The three constructions in the judgment experiment that test pronouns assigned both structural nominative and structural accusative case (numeral, PP, and NP modifiers) exhibit a larger difference in the relative acceptability of nominatives and accusatives in object position than in subject position (Table 15). This larger contrast in object position is to be expected if default case exerts a stronger effect on the grammar than prescriptivism. If one assumes that structural case, default case, and prescriptivism are implemented as OT-style constraints (Frazier 2007:14-15), an accusative in finite subject position would violate the assignment of nominative structural case, the most important constraint, but it would satisfy the default case constraint of general accusative case, arguably the second most important constraint. A nominative in object position similarly violates the assignment of structural case (accusative in these environments), but it also violates default case. This lesser satisfaction of highly ranked constraints makes nominative objects less “optimal” in a sense than accusative subjects. Additionally, nominative subjects are arguably less optimal than accusative objects; they both satisfy their respective structural case constraints, but nominatives can never satisfy the default case constraint. I believe that both of these differences result in the contrast between cases being smaller in finite subject position. Prescriptivism is still a relevant factor; its role will be made explicit in the more precise analysis of section 5.3. The work done for this paper makes the case that grammaticality exists on a gradient (rather than two discreet points) (Wasow 2009), and candidates are ranked by their overall optimality, variation resulting from speakers accessing and making use of less optimal candidates (Coetzee 2006).

5.2.2 THE SPECIAL STATUS OF *I*. Another general trend that emerges from the data from the judgment experiment is that *I* always received higher ratings than other nominatives where tested, both with the average raw and z-scores. The contrast is especially pronounced after *not*, where *I* received average ratings at least 2.4 points higher than the nearest nominative (nearly 3 for After *not* – Subject). *Me* also often received higher ratings than other accusatives, but the gap between *me* and the nearest accusative was only greater than 1 point in one case (In isolation – Object).

There are at least two possible explanations for the higher ratings given to *I*. The first is that the 1st person singular is subject to the most prescriptive pressure in English. Within COCA, *I* is the most frequently occurring pronoun (Table 30), and frequency certainly plays a role in the extension of case patterns subject to prescriptivism (Grano 2006:44). I would argue that evidence from hypercorrection (section 2.4), the marked formality of nominative pronouns in a number of environments (2.4), and the higher ratio of nominatives in more formal genres (4.3) indicate that prescriptivism in English is a force that increases the occurrence of nominatives (except in the case of *who/whom*) regardless of environment. This would interact with the high frequency of *I* such that *I* and *me* are under the most prescriptive pressure in English. If this is the case, then it is not surprising that *I* is consistently the most acceptable nominative in the environments tested.

Pronoun	# of tokens	Pronoun	# of tokens	Pronoun	# of tokens
<i>I</i>	5,362,818	<i>It</i>	5,123,568	<i>Him</i>	887,349
<i>You</i>	4,115,510	<i>We</i>	2,405,274	<i>Her</i>	533,289
<i>He</i>	3,825,631	<i>They</i>	2,447,344	<i>Us</i>	461,878
<i>She</i>	2,002,518	<i>Me</i>	939,302	<i>Them</i>	886,994

Table 30: Pronoun tokens on COCA

The second possible explanation is that many participants in the judgment experiment found 1st singular *I* and *me* to be more semantically felicitous or *natural* than other person/number combinations. As answers to a question like *Who took the cookies?*, *I* and *me* are perfectly natural, and this is indeed the metric on which participants were asked to rate the stimuli. This claim is hard to verify, but it may help explain why both *I* and *me* generally received higher ratings.

Despite strong evidence from the judgment experiment for an “*I*-rule”, this idea was not borne out in the corpus study. If *I* is consistently rated more acceptable than other nominatives, then it should constitute a larger portion of its person/number combination than other nominatives, especially in environments subject to prescriptivism, but this is not the case. In neither environment was the percentage NOM of 1st singular pronouns greater than for all the others. After *not* (Table 28), the percentage NOM of 1st singular pronouns was actually the *smallest*. This fact requires an explanation. There may be a divide between judgment and usage regarding pronoun case in English, but not much else can be said without further research.

5.3. A PROPOSED GRAMMAR FOR ENGLISH PRONOUN CASE. This section provides a possible theoretical analysis of English pronoun case preferences and variation.

The major generalization to emerge from the work done for this paper is that, prescriptivism aside, English pronoun case forms only display high variability when the DP is complex, which I am defining here as having more than one terminal node. In laymen’s terms, this could be defined as a DP that consists of more than one word. This generalization was first noticed to my knowledge by Arnold Zwicky and discussed in works such as Zwicky (2004), so I believe that it is appropriate to give him credit. The formalization given in (29), however, is in my terms.

(29) *Zwicky’s Generalization*

Case assignment/checking in English does not reliably penetrate DPs with more than one terminal node.

Particular cases are likely still assigned to DPs as a whole. Under my interpretation, a DP like *us linguists* is assigned structural nominative case in finite subject position and structural accusative case in object position, but this assignment does not reliably penetrate it. Structural case still exerts an influence, but all the case constraints actively compete with one another in complex DPs. Different variants can appear either via consciously re-ranking the case constraints in specific environments subject to prescriptivism or more generally from accessing less optimal forms.

Excluding PRO and possessive pronouns, which I discuss in my thesis, the constraints that I postulate to be active in the determination of pronoun case in English are given in (30)-(33). *SC*, *DC*, and *PC* stand for structural case, default case, and prescriptive case respectively.

- (30) **SC-NOM Spec-TP_{+FIN}** – Assign one violation to a DP that does not exhibit structural nominative case when assigned it. This applies (likely exclusively) to the subject of a finite clause.
- (31) **SC-ACC Spec-TP_{FIN}/Spec-vP/Spec-pP** – Assign one violation to a DP that does not exhibit structural accusative case when assigned it. This applies at least to the subjects of non-finite clauses (CP_{FOR}) and the complements of active verbs and prepositions.
- (32) **DC-ACC** – Assign one violation to a DP that does not exhibit default accusative case. This applies in all environments and takes precedence where no structural case is assigned, such as with pronouns in isolation, after *not*, and in the foci of *it*-clefts.

- (33) **PC-NOM** – Assign one violation to a DP that does not exhibit prescriptive nominative case. This applies in all environments, but its strength is highly variable.

The two structural case constraints are equally the most important ones. They apply nearly categorically to entire-DP pronouns that receive structural case, and they apply in mutually exclusive environments, making it difficult to rank one over the other. The default case constraint is the second most important. It is suppressed (though still active) when structural case is assigned, but its effects are clearly seen in environments where DPs do not receive structural case. For prescriptive case, the corpus study showed that prescriptivism increases the probability of nominative complements of *than* and *not*. Huddleston and Pullum (2002) discuss this increase in many other environments. The strength of prescriptive case is the most difficult to quantify, as it varies by sociolinguistic genre and syntactic environment. In general, evidence discussed in this paper indicates that explicitly taught, extralinguistic rules matter the least in the determination of pronoun case, so prescriptive case is usually the weakest constraint. Because of the inherently extralinguistic nature of prescriptive case, I would argue that nominative emerges in academic writing and other formal genres from a re-ranking of the constraints to prioritize it.

In order to illustrate how these constraints lead to the selection of different pronoun case forms, some example tableaux are presented below as Tableaux (1)-(4). The numbers preceding the candidates represent their rank in terms of optimality.

Position: Spec-TP _{+FIN}	SC-NOM ~ SC-ACC	DC-ACC	PC-NOM
✓ 1. We Californians		NA	*
2. Us Californians	*!	NA	*

Tableau 1: Subject of a finite clause

As the subject of a finite clause, a DP is assigned structural nominative case. Only *we Californians* satisfies this constraint, so it is the most optimal. *Us Californians* does not satisfy SC, but it satisfies DC and so is the second most optimal. Zwicky's Generalization allows both candidates to emerge in attested speech, as both are sufficiently optimal to do so.

Position: Spec-vP	SC-NOM ~ SC-ACC	DC-ACC	PC-NOM
2. We Californians	NA	*!	*
✓ 1. Us Californians	NA		*

Tableau 2: Object of an active verb

As the object of an active verb, a DP is assigned structural accusative case. Only *us Californians* satisfies SC, so it is the most optimal. The difference in the optimality of nominative and accusative pronouns is greater in this environment than in finite subject position, because the preferred *us Californians* also satisfies DC, but the dispreferred *we Californians* satisfies neither constraint. *We Californians* does satisfy PC, but this constraint is the weakest. This contrast leads to the correspondingly larger difference in ratings between nominatives and accusatives in object position. Once again, Zwicky's Generalization allows both candidates to emerge in attested speech, but nominatives are less possible here than accusatives in subject position.

Position: Ø	SC-NOM ~ SC-ACC	DC-ACC	PC-NOM
2. I	NA	NA	*!
✓ 1. Me	NA	NA	*

Tableau 3: In isolation

At least in English, pronouns in isolation do not receive structural case, so no SC constraint applies. Thus, default case prevails, and the pronoun receives default accusative case. These two DPs do not contain more than one terminal node, so accusative case reliably penetrates to the pronoun, making nominative *I*, the less optimal candidate, highly unlikely to surface.

Position: Spec-pP	PC-NOM	SC-NOM ~ SC-ACC	DC-ACC
✓ 1. Than I		NA	*
2. Than me	*!	NA	

Tableau 4: Object of *than* with prescriptive influence

In a more formal register like academic writing, nominative case can emerge from a re-ranking of the case constraints that prioritizes prescriptive case. For an object of *than*, I would argue that nominative case is an effect of prescriptivism, rather than the result of accessing a less optimal candidate, and that the conscious thought associated with prescriptivism causes this conscious re-ranking. If this is the case, then the only way that a single-terminal node nominative can regularly emerge in an environment in which both SC and DC favor the accusative is if PC is ranked above them. The environments in which this type of re-ranking can occur differ across and within speakers, so it is not easy to predict when it will apply, but I would say that this occurs most often in formal registers in environments such as *than* and *as* comparatives, coordinates, and any others where speakers are often corrected for using accusative pronouns.

6. Conclusion. In this paper, I have provided a review of the general principles surrounding case and their applications to English; I have described and presented the results of an acceptability judgment experiment conducted to determine the case preferences in the non-coordinated environments that exhibit the most variation (possibly the first of its kind on speakers of American English); I have presented the results of a corpus study performed to measure the influence of prescriptivism on English pronoun case; and perhaps most importantly, I have formulated an OT-style constraint grammar of English pronoun case that I believe provides an intuitive model of both preferences and variation. It is especially this last contribution that I hope will further linguists' understanding of case in natural language, especially in case-poor languages like English.

This paper has brought up a couple of potential lines of research that merit further inquiry. The first is a better understanding of the proposed *I*-rule for case. The judgment experiment provided evidence for it, but no such evidence came from the corpus study. More research would hopefully shed some light on this discrepancy between acceptability judgments and actual usage.

The second potential line of research is finding a theoretical explanation for Zwicky's Generalization, which captures an empirical observation about English but begs the question of *why* case assignment does not reliably penetrate multi-word DPs when it does so nearly categorically for entire-DP pronouns. I suspect that this fact of English has something to do with its impoverished and unproductive case system, but this is only speculation, and it makes one wonder about the typology of case variation and the precise mechanisms by which DPs receive case. It may be helpful to see if Zwicky's Generalization has broader applicability by examining the distribution of case forms in other languages, especially those that only have vestigial case marking on lexical items like pronouns, such as most of the modern Romance languages.

If this paper has demonstrated anything, it is that multiple factors are at play in the determination of case forms. Arriving at a full understanding of the intricacies of case in natural language will certainly not be easy, but I am optimistic and excited that over time we will gain a more complete understanding of this complex and endlessly interesting phenomenon.

References

- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: The MIT Press.
- Coetzee, Andries W. 2006. Variation as accessing ‘non-optimal’ candidates. *Phonology* 23(3). 337-385. <https://dx.doi.org/10.1017/S0952675706000984>.
- Frazier, Melissa. 2007. Default case in OT syntax. Ms. *Rutgers Optimality Archive* [ROA] 941. <http://roa.rutgers.edu/files/941-1207/941-FRAZIER-0-0.PDF>.
- Grano, Thomas. 2006. “Me and her” meets “he and I”: Case, person, and linear ordering in English coordinated pronouns. Stanford, CA: Stanford University undergraduate honors thesis. <http://pages.iu.edu/~tgrano/uht.pdf>.
- Hankamer, Jorge. 1973. Why there are two than’s in English. In Claudia Corum, T. Cedric Smith-Stark & Ann Weiser (eds.), *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society* [CLS 9]. 179-191. Chicago, IL: Chicago Linguistics Society.
- Hornstein, Norbert, Jairo Nunes & Kleanthes K. Grohmann. 2005. *Understanding minimalism*. Cambridge, UK: Cambridge University Press.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge, UK: Cambridge University Press.
- Martin, Roger. 2001. Null case and the distribution of PRO. *Linguistic Inquiry* 32(1). 141-166. <https://dx.doi.org/10.1162/002438901554612>.
- Parrott, Jeffrey K. 2006. Distributed Morphological mechanisms of pronoun-case variation. *University of Pennsylvania Working Papers in Linguistics* 12(2). 173-187. <http://repository.upenn.edu/cgi/viewcontent.cgi?article=1687&context=pwpl>.
- Quinn, Heidi. 2002. *The distribution of pronoun case forms in English*. Christchurch, NZ: University of Canterbury dissertation. <http://hdl.handle.net/10092/4850>.
- Schütze, Carson T. 2001. On the nature of default case. *Syntax* 4(3). 205-238. <https://dx.doi.org/10.1111/1467-9612.00044>.
- Sobin, Nicholas. 1997. Agreement, default rules, and grammatical viruses. *Linguistic Inquiry* 28(2). 318-343. <https://www.jstor.org/stable/4178979>.
- Wasow, Thomas. 2009. Gradient data and gradient grammars. In Malcolm Elliot, James Kirby, Osamu Sawada, Eleni Staraki & Suwon Yoon (eds.), *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society* [CLS 43]. 255-271. Chicago, IL: Chicago Linguistics Society. http://www.stanford.edu/~wasow/Wasow_CLS.pdf.
- Woolford, Ellen. 2006. Lexical case, inherent case, and argument structure. *Linguistic Inquiry* 37(1). 111-130. <https://dx.doi.org/10.1162/002438906775321175>.
- Zaenen, Annie, Joan Maling & Höskuldur Thráinsson. 1985. Case and grammatical functions: The Icelandic passive. *Natural Language and Linguistic Theory* 3(4), 441-483. <https://dx.doi.org/10.1007/BF00133285>.
- Zwicky, Arnold. 2004. Here comes the accusative. Post on *Language Log*. <http://itre.cis.upenn.edu/~myl/language-log/archives/001762.html>.
- Zwicky, Arnold. 2012. Dinosaur grammar. Post on *Arnold Zwicky’s Blog: A blog mostly about language*. <https://arnoldzwicky.org/2012/01/23/dinosaur-grammar/>.