# 2 Grammatical and Syntactic Variation

This chapter reviews the literature on grammatical variables (a.k.a. "alternations"), including from the perspective of traditional dialectology, and of modern variationist linguistics. This leads us to a brief review of well-known grammatical variables/alternations in English, a survey of previous comparative investigations of grammatical alternations in English, and finally to the presentation of the three grammatical (and more specifically, syntactic) alternations that will take center stage in the empirical chapters of the book: the genitive alternation, the dative alternation, and the particle placement alternation.

#### 2.1 Does Grammatical Variation Even Exist?

In this book we are interested in what van Hout and Muysken (2016, 250) call Type 3 variability, that is, "[v]ariability in the linguistic signal within a given language" - or, in Labovian parlance, "alternate ways of saying 'the same' thing" (Labov, 1972, 188). Now, whether or not that kind of variability exists in the realm of grammar (a term that in what follows we use synonymously with morphosyntax) has been controversial for a long while. Consider that the study of variation (and change) in pronunciation has a long history in linguistics, going back to seminal work on sound changes in (proto)languages such as Indo-European and Germanic (e.g. Grimm's law and Verner, 1877). Lexical variation has likewise concerned linguists for a long while. By contrast, grammatical variation in the sense that we are interested in – about choices languages users have between different grammatical constructions to express the same meaning or function – began receiving interest considerably later. There are several reasons why. For one thing, grammatical variation is arguably more abstract and thus less obvious to most people than pronunciational or lexical variation. Secondly, historically speaking there has been a gut feeling among some or even many descriptive linguists that grammatical variation - plain and simple - does not or should not exist, and that if it does exist it should be short-lived diachronically. In variationist circles, this gut feeling is

We hasten to add that prescriptive linguistics and language mavens have been commenting on variation between "good" and "bad" ways of saying the same thing for centuries, also and

also known as the "Doctrine of Form-Function Symmetry" (see Poplack, 2018, 7 for discussion). Take the well-known Principle of Isomorphism, formulated by Haiman citing authors such as Bloomfield and Bolinger, among others:

... the commonly accepted axiom that no true synonyms exist, i.e. that different forms must have different meaning... (Haiman, 1980, 516)

This principle simply denies that different ways, lexical or grammatical,<sup>2</sup> of saying the same thing exist. The same idea fuels the "Principle of No Synonymy" in Construction Grammar à la Adele Goldberg:

If two constructions are syntactically distinct, they must be semantically or pragmatically distinct. (Goldberg, 1995, 67)

Thus again the "Principle of No Synonymy" precludes the existence of different grammatical or syntactic ways of saying the same thing, and its impact in Construction Grammar and Cognitive Linguistics circles should not be underestimated.<sup>3</sup>

What is more, axioms such as the Principle of Isomorphism and the Principle of No Synonymy have also shaped thinking in historical linguistics, as De Smet et al. critically observe:

The relation between functionally similar forms is often described in terms of competition. This leads to the expectation that over time only one form can survive (substitution) or each form must find its unique niche in functional space (differentiation). (De Smet et al., 2018, 197)

However, as De Smet et al. (2018, 201) point out, "[i]f isomorphism were the only force shaping communicative codes, synonymy... should be consistently rooted out in diachrony." But it is not the case that grammatical variation phenomena are necessarily short-lived diachronically: for example, all three alternations subject to study in the present book have been around for centuries (see Section 2.5 for discussion). But the Doctrine of Form–Function Symmetry appears to be on shaky empirical grounds also in synchrony: the underlying assumption that believers in form–function symmetry often (implicitly) make is that grammatical variation is suboptimal, (needlessly) complex, cognitively costly, and therefore inconvenient for language users. Against this backdrop

particularly in the realm of grammar (see e.g. Anderwald, 2016 for discussion). We stress that when we diagnose a certain reluctance to accept the existence of grammatical variation, we exclusively refer to work in descriptive/theoretical linguistics.

- <sup>2</sup> Haiman clarifies that he is talking about "one-to-one correspondence[s] between the signans and the signatum, whether this be a single word or a grammatical construction" (Haiman, 1980, 515).
- Onsider, for example, that even recent work about grammatical variation published in journals such as Cognitive Linguistics regularly includes disclaimers such as the following: "Strictly speaking, there are no so-called syntactic alternations in natural languages. The semantics of alternative forms is never fully equivalent (Goldberg, 2002)..." (Fang and Liu, 2021, 2, in a paper about syntactic variation in Mandarin Chinese).

Gardner et al. (2021) investigate corpus data to check whether variable grammatical patterns attract production difficulties in the form of disfluencies (i.e. filled or unfilled pauses), as variable patterns must if the Doctrine of Form-Function Symmetry were to have a cognitive basis. However, Gardner et al. find no evidence whatsoever of a relationship between grammatical variation and disfluency. Therefore, it does not appear that grammatical variation is difficult to handle for language users.

For us the question of whether or not grammatical variation exists is an empirical one – and the empirical record, which includes a massive body of literature in grammatically oriented variationist (socio)linguistics and neighboring fields, strongly suggests that there is, in fact, "[Form-Function] asymmetry in the form of robust variability subject to regular conditioning" (Poplack, 2018, 7).

### 2.2 From Traditional Dialectology to Variationist Linguistics

Reservations about grammatical variation of the type discussed in the previous section have led to the situation that even fields specializing in analyzing variation, such as dialectology and variationist sociolinguistics, have traditionally prioritized variation on other linguistic levels than grammar.

Early dialectology was theoretically inspired by the Neogrammarian theorem of the "exceptionlessness of sound change" (Ausnahmslosigkeit der Lautgesetze – see Osthoff and Brugmann, 1878, XII–XIII), and so it is maybe not surprising that traditional dialect studies have relied mostly on questionnaires to elicit data about lexis and – in particular – pronunciation. Consider, for example, the Survey of English Dialects (SED) (Orton and Dieth, 1962): in this atlas project, only a small proportion of questions is dealing with morphosyntactic variation. The fact of the matter is that "modern dialectology has until recently largely neglected the field of syntax" (Görlach, 1999, 493). As Kortmann wrote in 2002, "Even today, when we look at current Anglo-American dialect research, there is no denying that the study of dialect syntax still constitutes no more than a sideline" (Kortmann, 2002, 187). The qualifier "Anglo-American" is important, because in the generative community the emergence of the Principles and Parameters framework in the 1980s triggered much interest in syntactic microvariation between regional varieties of the same language (Brandner, 2012); the bulk of the relevant literature deals with Italian, Dutch and Flemish dialects.

For non-generative dialectological circles, to the extent that the existence of grammatical variation is acknowledged, there continues to be a sense that grammatical variation exhibits relatively weak geographical patterns compared to other levels of language structure. For example, Lass (2004, 374) writes that "English regional phonology and lexis... are generally more salient and defining than regional morphosyntax" (see also Wolfram and Schilling-Estes, 1998, 161 and Löffler, 2003, 116).

With all that being said, we also need to keep in mind that dialectologists tend to be primarily interested in Type 4 variability according to the categorization in van Hout and Muysken (2016, 250), that is, "inter-individual variability": what are the forms that individuals in location X use, as opposed to those forms that individuals in location Y use? By contrast, Type 3 variability ("[v]ariability in the linguistic signal within a given language," van Hout and Muysken, 2016, 250) – or, in Labovian parlance, variability between "alternate ways of saying 'the same' thing" (Labov, 1972, 188) – is not something that traditional dialectologists much care about.

The discipline in variation studies that does care both about Type 3 and Type 4 variability, is variationist sociolinguistics. Early work in variationist sociolinguistics was often focused on pronunciational variation. In the 1970s analysts started to take more interest in grammatical variation as well (see e.g. Weiner and Labov, 1983 for early work on the active/passive alternation). This sparked a debate in the variationist sociolinguistics community as to whether linguistic variables can be grammatical in nature: Lavandera (1978) argued that the concept of the linguistic variable cannot be easily extended to linguistic levels other than phonology because "difficulties arise from the fact that non-phonological variation involves referential meaning" (p. 181); Labov (1978) replied that while it may be hard to work out the envelope of variation, it is in principle not impossible to extend the linguistic variable to levels such as grammar. The consensus that emerged after this debate is, then, that the concept of the linguistic variable can indeed be applied to linguistic levels other than phonology because "distinctions in referential value or grammatical function among different surface forms can be neutralized in discourse" (Sankoff, 1988, 153). Work on grammatical variation has boomed especially in Canada (see e.g. Poplack and Dion, 2009; Tagliamonte, 2014) and in the United Kingdom (see e.g. the papers in Beaman et al., 2021) and in variationist sociolinguistic work on languages such as Spanish (see e.g. Travis and Torres Cacoullos, 2012). But the fact remains that phonology is going strong in the variationist sociolinguistics community – certainly stronger than grammar.

By contrast, grammatical variation has been a focus from early on in corpus-based variationist linguistics (see e.g. Gries, 2005; Bresnan et al., 2007; Grondelaers and Speelman, 2007 for seminal work). In that community, grammatical "variables" (in variationist sociolinguistics parlance) are typically known as grammatical "alternations," defined as "structurally and/or lexically different ways to say functionally very similar things" (Gries, 2017, 7).

## 2.3 Grammatical Variation in English

Concerns about variation in the domain of grammar notwithstanding, the upshot is that grammatical variation in languages such as English is now quite

well documented and understood. In nonstandard English, well-known grammatical variation phenomena include the following (many of which are also covered in surveys such as Kortmann and Szmrecsanyi, 2004 and Kortmann and Lunkenheimer, 2013):

- copula retention versus deletion (as in *she is the first one* vs. *she the first one*; Labov, 1969)
- standard versus nonstandard reflexives (as in *they didn't go themselves* vs. *they didn't go theirself*; Hernández, 2010)
- demonstrative *those* versus demonstrative *them* (as in *in those days* vs. *in them days*; Britain, 2010)
- be versus have as perfect auiliaries (as in I'm come down to pay the rent vs. I've come down to pay the rent; Tagliamonte, 2000)
- *a*-prefixing on *-ing*-forms (as in *he was a-waiting there* versus *he was a-waiting there*; Wolfram, 1976)
- nonstandard verb forms (as in *he knowed* versus *he knew*; Szmrecsanyi, 2013a, features 28–30)
- standard negators versus ain't (as in people haven't got no money vs. people ain't got no money; Anderwald, 2003)
- multiple negation versus standard negation (as in *people haven't got no money* vs. *people haven't got any money*; Anderwald, 2005)
- the was—weren't split (as in It wasn't very dead, no it were just busy; Britain, 2002, 19)
- was-were variation (as in Three of them was killed vs. three of them were killed; Tagliamonte, 2009)
- standard versus non-standard verbal -s (as in I say vs. I says; Britain, 2010)
- the *kena*-passive in SgE (as in *John kena scolded* vs. *John was scolded*; Bao and Wee, 1999)
- unstressed preverbal *did* (as in *I did always eat* vs. *I always ate*; Jones and Tagliamonte, 2004).

We now know that there is likewise plenty of grammatical variation in more acrolectal, standard English of the type that is investigated in this book. Gardner et al. (2021), in their comprehensive study of the (nonexisting) relationship between grammatical variation and disfluency in mainstream US American English, study the following catalogue of grammatical variables/alternations (note that a corresponding list for British English or other varieties of English would look fairly similar):

- indefinite pronouns (as in *everyone would know* vs. *everybody would know*; D'Arcy et al., 2013)
- case and order of coordinated pronouns (as in *my husband and I vs. me and my husband*; Angermeyer and Singler, 2003)

- that versus zero complementation (as in I don't think it's a deterrent at all vs. I don't think that it's a deterrent at all; Tagliamonte and Smith, 2005)
- infinitival versus gerundial complementation (as in *I love to play racquetball* vs. *I love playing racquetball*; Mair, 2003)
- complementation after *remember*, *regret*, and *deny* (as in *I don't remember being that picky* vs. *I don't remember that I was that picky*; Cuyckens et al., 2014)
- complementation after the verb to try (as in *I try to get you some popcorn* vs. *I try and get you some popcorn*; Brook and Tagliamonte, 2016)
- particle placement (as in to pay off their credit cards vs. to pay their credit cards off; Gries, 2003)
- the dative alternation (as in *someone's given me one* vs. *someone's given one to me*; Bresnan et al., 2007)
- the genitive alternation (as in *they took someone else's life* vs. *they took the life of someone else*; Hinrichs and Szmrecsanyi, 2007)
- relativization (as in the house that you have to have money for vs. the house which you have to have money for; Hinrichs et al., 2015)
- analytic versus synthetic comparatives (as in *I think in Dallas it's a lot more scary* vs. *I think in Dallas it's a lot scarier*; Hilpert, 2008)
- plural existentials (as in *there's some places* vs. *there are some places*; Chambers, 2004)
- future temporal reference (as in *eventually it will happen* vs. *eventually it's going to happen*; Torres Cacoullos and Walker, 2009)
- deontic modality (as in *I must admit* vs. *I got to admit*; Tagliamonte and Smith, 2006)
- stative possession (as in *you got a Louisiana accent* vs. *you have got a Louisiana accent*; Tagliamonte et al., 2010)
- quotatives (as in *I was like, "No way!" and she goes, "Yeah!"*; Gardner et al., 2020)
- not versus no negation (as in There's nobody to sit them down vs. There isn't anybody to sit them down; Childs, 2017)
- negative versus auxiliary contraction (as in *he wouldn't admit it to you* vs. *he'd not admit it to you*; MacKenzie, 2013).

We add that, for practical reasons, Gardner et al. (2021) do not include grammatical alternations for which potentially every utterance would constitute a variable context. Such alternations include the active/passive alternation (as in they broke into the liquor closet vs. the liquor closet was broken into; Weiner and Labov, 1983), the progressive alternation (as in I love it vs. I'm loving it; Hundt, 2004), and variation between omitted versus overt pronominal subjects (as in Tom says hi. Asks how you are doing vs. Tom says hi. He asks how you are doing; Torres Cacoullos and Travis, 2014). Other grammatical alternations in standard English that have been studied elsewhere include would versus

used to as markers of habitual past (as in *I used to dance* vs. *I would dance*; Tagliamonte, 2000), and the mandative subjunctive alternation (as in *the law requires that the doors be shut* vs. *the law requires that the doors must be shut*; Kastronic and Poplack, 2014).

We conclude our survey of grammatical variation in English by noting that grammatical variables/alternations can be categorized into three general types (see De Troij, in preparation):

**Permutation alternations** Permutation alternations allow language users to manipulate constituent ordering. Consider (4). In (4-a), the pronoun follows the full NP while in (4-b), it precedes; constituent order is variable.

### (4) The case and order of coordinated pronouns alternation:

- a. ... you give Al Gore and I a chance to bring America back. (Angermeyer and Singler, 2003, ex. (1a))
   (NP + pronoun order)
- b. in the interchanges between *she and Chairman Fazio* ... (Angermeyer and Singler, 2003, ex. (1b)) (pronoun + NP order)
- **Insertion/deletion alternations.** These alternations give language users the option of retaining or omitting functional markers or patterns. Consider (5). In (5-a), the complementizer *that* is omitted; in (5-b), it is retained.

### (5) The complementizer retention/omission alternation:

- a. To prove Ø I could do it.... (Tagliamonte and Smith, 2005, ex. (1)) (complementizer omission)
- b. ... Yes, I had to prove *that* I could do it. (Tagliamonte and Smith, 2005, ex. (1)) (complementizer retention)
- **Substitution alternations.** Substitution alternations replace one particular function word or functional construction by another function word or functional construction. Consider (6). In (6-a), the speaker uses the future marker *will* instead of *be going to*; vice versa for (6-a).

### (6) The future marker alternation:

a. And he'll probably live 'til a hundred. (Torres Cacoullos and Walker, 2009, ex. (1c))

(the future marker *will*)

b. My doctor tells me I'm going to live 'til a hundred. (Torres Cacoullos and Walker, 2009, ex. (1d)) (the future marker be going to)

### 2.4 Comparative Perspectives on Grammatical Variation in English

In this section, we review previous comparative work on grammatical variation in English (see Section 1.2): how do grammatical choice-making processes in location/variety/dialect A differ from those in location/variety/dialect B?

The field of comparative sociolinguistics is too large to review exhaustively here, therefore our discussion will necessarily be somewhat selective. As to comparisons involving vernacular dialects in North America, for example, in a seminal study Poplack and Tagliamonte (1989) examine variation in verbal -s inflection in two corpora on early Black English from a historical and comparative perspective; Walker (2007) studies variable agreement in existentials with plural reference in varieties of Quebec English; and Gardner (2017) uses the comparative method across three grammatical variables to asses the genetic similarity between Cape Breton English (Canada) and its two potential progenitors, Scottish English and United Empire Loyalist English. As to comparisons involving vernacular dialects in the United Kingdom, Tagliamonte and Smith (2002) investigate negative/auxiliary contraction in eight communities (Tiverton, Henfield, York, Wheatley Hill, Maryport, Cullybackey, Cumnock, Buckie); Jones and Tagliamonte (2004) compare variation patterns between unstressed preverbal did and plain verb forms in Somerset to those in Samaná in the Dominican Republic, showing that the similarities are due to diffusion; and Childs (2017) studies variation between negative constructions in three UK locations (Glasgow, Tyneside, Salford). Beyond the United Kingdom and North America, we would like to mention Tagliamonte et al. (2016a), who study quotative variation in four speech communities around the world: Toronto (Canada), Victoria (Canada), Christchurch (New Zealand), and Perth (Australia).

In less sociolinguistically and more theoretically oriented circles, analysts interested in the probabilistic nature of knowledge about grammar and grammatical variation have recently also taken an interest in the comparative analysis of variation data. In this spirit, Bresnan and Hay (2008) explore the dative alternation in US American English (based on the Switchboard Corpus) and New Zealand English (based on the Origins of New Zealand English corpora) and find that a US regression model generalizes remarkably well to the New Zealand data, although New Zealand English speakers are more sensitive to the animacy constraint. Szmrecsanyi et al. (2017) investigate two alternations – the dative as well as the genitive alternation – in four varieties of spoken English: US American English (based on materials from the Switchboard Corpus and the Corpus of Spoken American English), British English (materials from the Freiburg English Dialect Corpus, and materials collected in the United Kingdom between 1997 and 2010 by Tagliamonte), Canadian English (materials from the sociolinguistic interviews in the Ontario Dialects Archive), and New Zealand English (materials from the Canterbury Corpus of the Origins of New Zealand English). The study concludes that "while there are a number of subtle probabilistic contrasts between the regional varieties under study, there is overall a striking degree of cross-varietal homogeneity" (Szmrecsanyi et al., 2017, 1).

Meanwhile, scholars interested in divergence/convergence patterns have endeavored to enrich comparative variation analysis with the addition of real time as an interacting language-external factor. Consider Hinrichs and Szmrecsanyi (2007), who study the genitive alternation in written-editedpublished corpus materials from the Brown family of corpora, contrasting both American English against British English and 1960s usage against 1990s usage; multivariate analysis indicates that the spread of the s-genitive in both British English and American English is primarily due to economy-related factors, while one of the reasons why the s-genitive is now more popular in American English than in British English is that the animacy constraint is weaker in American English than in British English. Hinrichs et al. (2015) investigate the same data source to study relativizer variation (which vs. that) in late twentieth-century written-edited-published English, likewise using comparative multivariate regression analysis. Results show that the welldocumented spread of relative that in restrictive relative clauses is what Hinrichs et al. (2015, 806) call a case of "colloquialization-cum-Americanization." Going a bit farther back in real time, Hundt and Szmrecsanyi (2012) investigate the progressive alternation and the genitive alternation in nineteenth- and early twentieth-century written New Zealand English (tapping into A Corpus of Early New Zealand English), as well as in British English and American English of the same period (tapping into A Representative Corpus of Historical English Registers – ARCHER); it is shown that there are differences between earlier New Zealand English and the other varieties in terms of the effect that animacy has on grammatical variation, similar to what Bresnan and Hay (2008) report. Finally, tapping into ARCHER also, Wolk et al. (2013) study the development of the dative and the genitive alternations during the late Modern English period (sixteenth to twentieth century). Their dative alternation materials include data from both British English and American English. It turns out that in American English, increasing theme length has come to disfavor the prepositional dative more robustly than in British English.

Lastly, within the World Englishes community comparative multivariate/probabilistic variation analysis is a relatively new addition to the analytical toolbox. Gries and Bernaisch (2016) model the dative alternation in six South Asian Englishes (English in Bangladesh, India, the Maldives, Nepal, Pakistan and Sri Lanka), based on the SAVE Corpus. Based on a regression-based technique – Multifactorial Prediction and Deviation Analysis with Regression (MuPDAR) – the study reveals that Indian English is the linguistic epicentre of South Asian English. In a similar spirit, Heller et al. (2017a) investigate the genitive alternation in six varieties of English (English in Great

Britain, Hong Kong, India, the Philippines, Singapore, and Sri Lanka) as represented in the International Corpus of English (ICE); based on a conditional random forest modeling-based approach (Multifactorial Prediction and Deviation Analysis Using Regression/Random Forests, or MuPDARF for short), Heller et al. argue, as Gries and Bernaisch (2016) do, that Indian English has epicenter status. Moving on, Levshina (2018) studies complementation patterns after the verb to help in seven varieties of English (English in Australia, Ghana, Great Britain, Hong Kong, India, Jamaica, and the United States) covered in the Corpus of Global Web-based English (GloWbE); what the study demonstrates via Bayesian regression is that universally, predictability of units in discourse is an important constraint on the variation, but that this being said "the strength, shape and directionality of predictability effects exhibit variation across the countries" (Levshina, 2018, 1). Davydova (2019) investigates quotative like in materials obtained from two student communities, one in Outer Circle India (Jawaharlal Nehru University), and the other in Expanding Circle Germany (University of Mannheim). Regression analysis shows that "the patterns of use of innovative be like are uniform in the speech of young adults with elevated levels of exposure to mass media products and apply across the Outer and Expanding Circle English board" (Davydova, 2019, 578). Finally, we would like to mention Tamaredo et al. (2019), who tap into the International Corpus of English (ICE) to explore the degree of regional variability concerning the probabilistic conditioning of four grammatical alternations in English (the genitive, dative, particle placement, and subject pronoun omission alternations) in British, Indian and Singapore English; Tamaredo et al. report that the genitive alternation is the most stable/homogeneous one across the three varieties, while the particle placement alternation is most amenable to probabilistic indigenization.

### 2.5 Grammatical Alternations Subject to Study in This Book

Of the many alternations in the grammar of English (see above), we investigate three: The genitive alternation, as in (1) (reprinted as (7) below); the dative alternation, as in (2) (reprinted as (8) below); and the particle placement alternation, as in (3) (reprinted as (9) below).

### (7) The genitive alternation in English:

a. Two other journalists who wrote a book criticising [the president] 
possessor's [brother]possessum were ordered to pay £6.3 million in fines (GloWbE AU B vexnews.com) 
(the s-genitive)

b. Can you imagine a couple of years after WW2 the allies permitting [the brother]<sub>possessum</sub> of [the president]<sub>possessor</sub> bankrupting the central bank through embezzlement and getting away with it? (GloWbE GB G guardian.co.uk) (the of-genitive)

### (8) The dative alternation in English:

- a. A victim will be asked to  $give_{verb}$  [the police] $_{recipient}$  [a statement]  $_{theme}$  explaining what has happened. (GloWbE CA G slsedmonton.com)
  - (the ditransitive dative)
- b. Neither of them gave<sub>verb</sub> [a statement]<sub>theme</sub> to [the police]<sub>recipient</sub>. (GloWbE JM G jamaicaobserver.com) (the prepositional dative)

### (9) The particle placement alternation in English:

- a. For all my second language readers: no need to look<sub>verb</sub> [the word]<sub>NP</sub> up<sub>particle</sub> in the dictionary... (GloWbE NZ B dedepuppets.com)
   (V-DO-P, a.k.a. the split variant)
- b. Look<sub>verb</sub> up<sub>particle</sub> [the word]<sub>NP</sub> in a dictionary and write down its meaning in a vocabulary notebook. (GloWbE US G artofmanliness.com)

(V-P-DO, a.k.a. the continuous variant)

All three alternations are first and foremost permutation alternations (see Section 2.3): by switching between variants, language users can change the order of possessor and possessum (genitive alternation), of recipient and theme (dative alternation), or of direct object and particle (particle placement alternation). Hence, the alternations studied here are not only grammatical but more specifically syntactic. That said, note that the genitive alternation can additionally be classified as a substitution alternation, in that language users can replace the genitive clitic -s by the preposition of. The dative alternation, on the other hand, can also be considered an insertion/deletion alternations, as the prepositional dative variant allows language users to retain the preposition to, while they can omit it if they use the ditransitive dative variant.

These alternations were chosen partly because the constraints governing each of them are well-known, and have been shown to exhibit regional variation in small-scale studies (see Bresnan and Hay, 2008; Wolk et al., 2013; Haddican and Johnson, 2012). However, the scope of variation on a larger scale is still poorly understood. Additionally, many constraints (animacy, phonology, semantic class, end-weight, information status, persistence) are shared across two or even all three alternations, which facilitates comparative analysis across alternations. The variation we find among these alternations is particularly illustrative of the often subtle, experienced-based nature of grammatical

knowledge, given that – to the best of our knowledge – none of these alternations have been shown to vary substantially along other social dimensions (e.g. age, gender, ethnicity, class), nor are they discussed as examples of regional indicators.

At the outset we would like to stress that per alternation, we use identical variable context definitions across the varieties of English under study. The rationale is twofold. First, we do not have reason to believe, and have not come across data suggesting that the contexts in which the syntactic variants are interchangeable differ substantially across the varieties we study. Second, the assumption of identical variable contexts is a prerequisite for comparative variationist analysis.

In the next sections we summarize the literature on each alternation.

### 2.5.1 The Genitive Alternation

The literature on the genitive alternation is vast; for a detailed literature review, we refer the reader to Rosenbach (2014) and Heller (2018). Variation between the *s*-genitive and the *of*-genitive is old. While the *s*-genitive goes back to the "properly" Germanic way of expressing genitive relations with inflectional means, the *of*-genitive is an innovation that appeared during the ninth century. Thomas (1931, 284) shows that frequency-wise, the inflected genitive greatly outnumbered the periphrasis with *of* up until the twelfth century. Then, starting in Middle English, we begin to see "a strong tendency to replace the inflectional genitive by periphrastic constructions, above all by periphrasis with the preposition of" (Mustanoja, 1960, I:70). In texts from the Early Modern English period, however, we observe a revival of the *s*-genitive, "against all odds" (Rosenbach, 2002, 184). The *s*-genitive has continued to be on the rise during the Late Modern English period (see e.g. Wolk et al., 2013; Szmrecsanyi et al., 2014; Hinrichs and Szmrecsanyi, 2007).

In Present-Day English s-genitive and of-genitive constructions at large can encode a "a grab-bag" (Givón, 1993, 264) of relationships between what is conventionally called the "possessor" NP (in (7), the president) and the "possessum" (in (7), brother). The point is that these labels do not necessarily mean that we are talking about possession: for the s-genitive alone, Quirk et al. (1985, 321–322) list eight different meanings: possessive, but also subjective, objective, the genitive of origin, descriptive, the genitive of measure, the genitive of attribute, and the partitive genitive. Crucially, not all of these can also be encoded by the of-genitive to the same extent, and vice versa. It is fair to say that there is no overwhelming consensus in the literature as to the importance of semantic distinctions: some analysts have claimed that possessive relations have a privileged status in the semantics of the s-genitive (Taylor, 1989), though others have argued that the two genitives convey generally the same meaning (e.g. Altenberg, 1982, 11), or – conversely – that the two genitive constructions are semantically empty (e.g. Hudson, 1984).

This difficulty in defining the meaning of genitive constructions (cf. Strang, 1968, 109: "any attempt to sum up 'the meaning' of the *s*-genitive is doomed") is why of the three alternations studied here, the genitive alternation is arguably the most tricky one when it comes to defining those contexts in which both variants are interchangeable: variationist analysis investigates variation between formally different but semantically/functionally equivalent constructions, but when the semantics of one or all variants is hard to pin down then the task becomes difficult. The way in which the recent variationist literature on the genitive alternations handles this difficulty is by distinguishing between choice contexts (i.e. contexts in which both variants can be used) and "knockout contexts" (see Rosenbach, 2014, Section 2.2.3 for detailed discussion), in which only one of the variants can be used. The trick here is to not define choice contexts directly, but to eliminate knock-out contexts until what is left is choice contexts only (see Chapter 4 for details).

In choice contexts, we see a range of known probabilistic constraints that regulate variation between genitive variants. One the one hand, we see in the genitive alternation factors that regulate permutation alternations in general: constituent animacy, constituent definiteness, information status (old before new), and constituent length/weight. On the other hand, the alternation is governed by constraints that are specific to genitive variation, chiefly among them the semantic relation between possessor and possessum, and the phonological shape of the possessor (final sibilancy). We will discuss the constraints considered in this book on page 26.

To conclude this section, we would like to mention in passing another complication: the vast majority of the literature restricts attention to binary variation between the *s*-genitive and the *of*-genitive. However, there is a third variant – the NN-genitive – which is sometimes interchangeable with either the *s*-genitive or the *of*-genitive or with both (as in *the FBI director*, which paraphrases both *the FBI's director* and *the director of the FBI* – see Szmrecsanyi et al., 2016b for a detailed analysis). In our analysis, we take the liberty to set aside the NN-genitive.

#### Previous Variationist Research on the Genitive Alternation

In this section we review recent variationist research on the genitive alternation. We restrict attention to studies that employ modern multivariate analysis techniques. Szmrecsanyi (2006, Chapter 5) (see also Szmrecsanyi and Hinrichs, 2008; Szmrecsanyi et al., 2017 for reanalyses of the dataset) is to the best of our knowledge the first multivariate study of the genitive alternation in English; investigating corpora sampling spoken British English and American English, the study uses regression analysis to show that previous usage of an s-genitive primes future usage of an s-genitive ( $\alpha$ -persistence), and that usage of the preposition of in non-genitive contexts primes usage of the of-genitive

 $(\beta$ -persistence). Hinrichs and Szmrecsanyi (2007) (see also Szmrecsanyi and Hinrichs, 2008; Szmrecsanyi, 2010 for reanalyses) investigate genitive variation in the Brown family of corpora (written-edited-published British English and American English, 1960s and 1990s) using regression analysis; their analvsis shows that the spread of the s-genitive in written English is primarily due to economy, given that the s-genitive is the more economical coding option. Hundt and Szmrecsanyi (2012) conduct a study of the genitive alternation in nineteenth- and early twentieth-century written British and New Zealand English (based on the ARCHER and CENZE corpora) and demonstrate via binary logistic regression analysis that possessor animacy is a more powerful constraint in the New Zealand English material compared to the British English material. Wolk et al. (2013) (see also Szmrecsanyi, 2013b; Szmrecsanyi et al., 2014; Ehret et al., 2014 for follow-up research) study both the dative and the genitive alternation in ARCHER, a corpus that covers materials from the Late Modern English period; the regression analysis in Wolk et al. (2013) shows that both the possessor animacy constraint and the possessum length constraints are diachronically unstable. The regression analysis in Grafmiller (2014), based on materials from the Switchboard Corpus and the original Brown Corpus, reveals substantial variation across six spoken and written genres concerning the effect sizes of language-internal constraints governing the choice between the s-genitive and the of-genitive in American English. Jankowski and Tagliamonte (2014) (based on Jankowski, 2013; see also Szmrecsanyi et al., 2017 for a re-analysis of the dataset) undertake a sociolinguistic analysis of genitive variation based on sociolinguistic interview materials from the Toronto English Archive (TEA) and the Southeastern Ontario Archive; Varbrul analysis indicates, among other things. that the animacy constraint is near-categorical in the materials, and that nonhuman collective/organization possessors increasingly attract the s-genitive in apparent time. Based on a dataset drawn from the Switchboard Corpus of US American English (used in Grafmiller et al., 2016 and reanalyzed in Szmrecsanyi et al., 2017), Shih et al. (2015) utilize regression analysis to show that genitive choice is not only constrained by animacy, weight, and other wellknown language internal factors but also by eurhythmicity: all other things being equal, language users use that genitive variant which maximizes the extent to which stressed and unstressed syllables alternate in an utterance. Heller et al. (2017a) apply Multifactorial Prediction and Deviation Analysis Using Regression/Random Forests (MuPDARF) to a dataset of genitive variation in materials from six components of the International Corpus of English (Great Britain, Hong Kong, India, the Philippines, Singapore and Sri Lanka); analysis shows that Indian English has epicenter status. Hackert and Wengler (2022) conduct a diachronic analysis of the genitive alternation in five varieties of English based on newspaper materials from the Bahamas, Jamaica, India, Great Britain, and the US; conditional random forest and

MuPDARF analysis shows that in particular Caribbean varieties are partaking in American-led global trends in grammar toward, for example, densification (via usage of the more compact *s*-genitive), without actually approximating American norms, while overall there is a widening gap between metropolitan and postcolonial Englishes. As to learner English, Dubois et al. (2022) regress genitive variants in the Trinity Lancaster Corpus and demonstrate that while native speakers and learners have remarkably similar probabilistic grammars, low-proficiency learners are less sensitive to possessor definiteness and possessor animacy than high-proficiency learners.

There is also substantial work on the genitive alternation in varieties of English by Benedikt Heller and colleagues (Heller et al., 2017b; Heller, 2018; Heller and Szmrecsanyi, 2019; see also Szmrecsanyi et al., 2016a; Tamaredo et al., 2019), which we summarize below and throughout this book.

#### Probabilistic Constraints on the Genitive Alternation

The genitive alternation is known to be sensitive to a number of probabilistic factors, the most important of which include the following:

- Animacy of the possessor is one of the most important constraints on genitive choice (see e.g. Rosenbach, 2005; Szmrecsanyi et al., 2017): animate possessors favor the *s*-genitive, while inanimate possessors disfavor.
- According the principle of end-weight (Behaghel, 1909; Wasow and Arnold, 2003), language users in VO-languages such as English have a preference for placing heavier constituents after less heavy constituents. This is why according to the literature decreasing possessor length favors the *s*-genitive (as the *s*-genitive places the possessor before the possessum) (see Rosenbach, 2014, table 2); the reverse pattern is true for the *of*-genitive.
- The literature suggests that possessors high on the definiteness scale attract the *s*-genitive, all other things being equal (see e.g. Rosenbach, 2014, table 2).
- Thanks to a haplology effect, the presence of a final sibilant in the possessor, as in *President Bush's speech*, is known to disfavor usage of the *s*-genitive (see e.g. Zwicky, 1987).
- Genitive choice is subject to priming effects (Szmrecsanyi, 2006; Hinrichs and Szmrecsanyi, 2007).
- Prototypical semantic relations have been shown to favor the *s*-genitive (see e.g. Rosenbach, 2014, table 2). Prototypical relations include body part relations (*Tom's hand*), kinship relations (*Tom's sister*), legal ownership (*Tom's house*), and part-whole relations (*the book's cover*).

More information on how we operationalized these factors is provided in Chapter 4.

#### 2.5.2 The Dative Alternation

The dative alternation, like the genitive alternation, is an extremely wellresearched syntactic variable; an extensive literature review is provided in Röthlisberger (2018a, Chapter 2). Prepositional dative constructions are attested as early as in Old English, though in that period the construction was subject to lexical restrictions (Allen, 2006, 206; De Cuypere, 2015). Additionally, constituent order in Old English ditransitive constructions was variable: the recipient could precede the theme, as in Modern Standard English, or it could follow the theme (van Kemenade, 1987; Koopman, 1990). More robust variability between the ditransitive dative and the prepositional dative dates back to the Middle English period (McFadden, 2002, 112), when the prepositional dative expanded into "a fully productive alternative" (Fischer and van der Wurff, 2006, 166) to the ditransitive dative. The emergence of the prepositional dative as a more analytic/periphrastic construction is sometimes speculated to have been triggered by the loss of case distinctions (see e.g. McFadden, 2002; Fischer and van der Wurff, 2006); an alternative explanation is contact with French (Visser, 1963). After Middle English, constituent order of nominal (but not pronominal) constituents – see Gerwin (2013) – in the ditransitive dative construction became fixed along the lines of the Modern English pattern by the late-fourteenth century (Allen, 2006, 206). In this book, we restrict attention to prepositional dative constructions in which the theme precedes the recipient, and ditransitive dative constructions in which the recipient precedes the theme.

The dative alternation has received attention in many subfields in linguistics, including e.g. by researchers interested in first/second language acquisition (e.g. Campbell and Tomasello, 2001), information status (e.g. Thompson, 1990), psycholinguistics (e.g. Bock, 1986), and collustructional preferences (e.g. Mukherjee and Gries, 2009). On a more general level, following Gerwin (2014, 19) the sizable body of literature on the dative alternation in English can be divided into two types: studies that adopt a single-meaning approach (popular in usage-based/variationist circles), as opposed to studies that adopt a multiple-meaning approach (which is dominant in the generative community, but also in Construction Grammar). The multiple-meaning approach is summarized by Bresnan et al. as follows:

Advanced by Green (1974) and Oehrle (1976) [references omitted], this idea [the multiple meaning approach] was taken up in influential work on language learnability by Pinker and colleagues (Gropen et al., 1989; Pinker, 1989). They argued that there are two ways of viewing the same giving event: as *causing a change of state (possession)* or *as causing a change of place (movement to a goal)*. They hypothesized that the different ways of conceptualizing the giving event are associated with different structures, the possession meaning with the double object structure and the movement meaning with the prepositional dative structure... (Bresnan et al., 2007, 71)

Subsequently Bresnan et al. present an extended empirical argument why the multiple-meaning approach is inadequate. Their summary is as follows:

First, linguistic intuitions of ungrammaticality are a poor guide to the space of grammatical possibility. Second, usage data reveals generalizations which we are sometimes blind to. Third, English dative verbs have more syntactic flexibility than we thought, occurring more freely in alternative constructions. And fourth, we cannot predict the dative alternation from meaning alone. (Bresnan et al., 2007, 75)

The alternative view is the single-meaning approach, whereby the ditransitive and prepositional dative constructions have essentially the same meaning, and variant choice is a function of language-internal (contextual) constraints such as constituent weight/length, constituent pronominality, information status, and so on. It is this variationist perspective that informs our analysis.

### Previous Variationist Research on the Dative Alternation

To our knowledge, the first study to use regression analysis to model the dative alternation is Williams (1994). The paper investigates a corpus of American English covering four spoken genres and eight written genres. The regression analysis tests a range of predictors: syntactic class of the verb, register, modality, givenness, prosodic length, definiteness, animacy, and specificity. Of these, prosodic length, syntactic class of verb, and register are identified as most important. Gries (2005) explores syntactic priming effects in the dative alternation (and also the particle placement alternation, based on the ICE-GB corpus). Regression analysis shows that there are significant priming effects comparable to those reported in the experimentalist literature on the dative alternation. That said, the strength of priming effects differs across dative verbs. Bresnan et al. (2007) present a regression model of the dative alternation in spoken American English sampled from the Switchboard Corpus. They demonstrate that the alternation is predictable from a series of language-internal probabilistic constraints: givenness of recipient/theme, pronominality of recipient/theme, definiteness of recipient/theme, animacy of recipient, person of recipient, number of recipient/theme, concreteness of theme, and syntactic parallelism (see Bresnan, 2007 and Bresnan and Ford, 2010 for follow-up investigations of the Bresnan et al., 2007 dataset adding an experimentalist twist via a rating task, and Szmrecsanyi et al., 2017 for a comparative re-analysis of the dataset). Bresnan and Hay (2008) compare US American and New Zealand English datives (see Section 2.4). Theijssen et al. (2013) is a methodological paper that investigates a dative alternation dataset extracted from the British National Corpus (BNC) using Bayesian Networks (Pearl et al., 1988), memory-based learning (Daelemans and van den Bosch, 2005), and logistic regression analysis. Theijssen et al. (2013) conclude from their multi-method approach that "most of the alternation is determined by the verb and the length of the two objects" (Theijssen et al., 2013, 227). De Cuypere and Verbeke (2013) analyze the dative alternation in a sample extracted from the Kolhapur corpus of Indian English. Regression analysis shows that dative verb, recipient pronominality and constituent length are significant predictors of dative choice in their

dataset. Using data from ARCHER (a corpus of materials from the Late Modern English period), Wolk et al. (2013) conduct a regression analysis of the dative alternation which indicates that the disfavoring effect of inanimate recipients towards the ditransitive dative variants weakened during the Late Modern English period. Results indicate that theme length has a stronger effect on variant choice in American English. Schilk et al. (2013) investigate variation between complementation patterns (including the ditransitive and prepositional dative variants) after the verb give in South Asian Englishes based on web-derived corpora; multinomial regression indicates that "Pakistani English is closer to British English with regard to relevant driving factors than Indian English" (Schilk et al., 2013, 187). Similarly, Bernaisch et al. (2014) explore the dative alternation in six East Asian varieties (Bangladeshi English, Indian English, Maldivian English, Nepali English, Pakistani English and Sri Lankan English) plus British English as the reference variety, based on materials from the South Asian Varieties of English (SAVE) Corpus and the BNC; conditional inference trees and conditional random forest modeling uncovers "variety-independent protostructions, that is, abstract combinations of crossvarietally stable features with high predictive power for a particular syntactic pattern" (Bernaisch et al., 2014, 7). A comparative sociolinguistics analysis of the dative analysis is presented in Tagliamonte (2014), based on vernacular spoken language data (sociolinguistic interviews) collected in the United Kingdom and in Canada (see Szmrecsanyi et al., 2017 for a reanalysis of the dataset); regression modeling and conditional inference tree analysis fails to uncover any major differences as to how the dative alternation works in the Canadian and UK materials. Engel et al. (2022) investigate the extent to which the effect of probabilistic constraints on the dative alternation in British English differs as a function of register. Analysis shows that register modulates the probabilistic effects of definiteness of the dative constituents.

What will take center stage in this book is work on the dative alternation by Melanie Röthlisberger and colleagues (Röthlisberger et al., 2017; Röthlisberger, 2018a; see also Szmrecsanyi et al., 2016a; Tamaredo et al., 2019), which we summarize below and throughout this book.

#### Probabilistic Constraints on the Dative Alternation

The following factors are the "usual suspects" in the literature on the dative alternation:

• The principle of end-weight (Behaghel, 1909; Wasow and Arnold, 2003) predicts that heavier themes should follow less heavy recipients, while heavier recipients should follow less heavy themes. Such end-weight effects are very well documented for the English dative alternation (see e.g. Bresnan et al., 2007; Szmrecsanyi et al., 2017).

- According to the literature, pronominal recipients strongly favor the ditransitive dative variant (see e.g. Bresnan et al., 2007; Szmrecsanyi et al., 2017).
- Bresnan and Ford (2010) show that definite dative constituents tend to be placed first.
- Information status is a well-known determinant of constituent order variation, including the dative alternation (see e.g. Bresnan et al., 2007): language users tend to place discourse-given constituents before discoursenew constituents if they have a choice (see e.g. Arnold et al., 2000; Collins, 1995).
- More animate recipients/themes tend to be placed before less animate constituents (again, see e.g. Bresnan et al., 2007; Szmrecsanyi et al., 2017).

More information on how we operationalized the above factors is provided in Chapter 4.

### 2.5.3 The Particle Placement Alternation

After many transitive particle verbs in English, language users have the choice between placing the particle after the direct object (as in look the word up – the "split" variant) or before the direct object (as in look up the word – the "continuous" variant). This syntactic variable we refer to as the particle placement alternation. While there is a consensus that the two variants are semantically equivalent, previous qualitative and nonvariationist quantitative scholarship on the particle placement alternation has sought to uncover the pragmatic and discourse-functional differences between the two constituent order patterns (see e.g. Fraser, 1965; Bolinger, 1971; Fraser, 1976; Gries, 1999; Dehé, 2002; Lohse et al., 2004; Cappelle, 2005, 2009; Thim, 2012). This literature, which inspires the constraints that we model, agrees that particle placement is a probabilistic phenomenon, with no single factor categorically determining the choice that language users have. Given that we also cover a number of indigenized L2 varieties, we should also add that phrasal verbs are known to pose a challenge to learners of English, especially those whose L1 does not have phrasal verb constructions (see e.g. Liao and Fukuya, 2004; Siyanova and Schmitt, 2007; Alejo-González, 2010; see also Gries, 2011 for a discussion of L1 acquisition of particle verbs).

The particle placement alternation in its modern form with postverbal variants is attested from the Middle English period onwards – in Old English, the particle could also be placed before the verb (Claridge, 2000, 85), as in modern Germanic languages such as Dutch or German. In the Middle English period, the preverbal variants disappeared (Hiltunen, 1983, 106–111; Thim, 2012, 103), leaving behind the variation between the two postverbal variants that we still see in Modern English. The conditioning of this variation in Middle

English was also similar to that in Modern English, in that the continuous variant was preferred with NP objects and the split variant with pronominal objects (Elenbaas, 2013). As to variant rates, analysts report a predominance of the continuous variant in Early, and Late-Modern English texts (see Elenbaas, 2013, 495; Rodríguez-Puente, 2016, 150), although substantial register differences exist.

Previous Variationist Research on the Particle Placement Alternation Gries (2003) (see also Gries, 2002 for a partial summary) is the first modern variationist analysis of particle placement. Gries investigates the British National Corpus to study the effect of a range of constraints discussed in the literature via discriminant analysis, a multifactorial technique. Results indicate that particle placement is primarily a function of discourse-functional factors. Gries (2005) explores syntactic priming effects on the particle placement alternation, beside the dative alternation (see Section 2.5.2), based on the ICE-GB corpus). Regression analysis uncovers significant priming effects comparable to those reported in the experimentalist literature while at the same time particle verbs differ in their amenability to priming effects. Priming effects (a.k.a. persistence effects) were also tested in Szmrecsanyi (2006 chapter 7) (see Szmrecsanyi, 2005 for a partial summary): here the particle placement study is based on materials from the Corpus of Spoken American English (CSAE) and the Freiburg Corpus of English dialects (FRED). A regression analysis indicates that usage of the split variant robustly decreases the odds that the continuous variant will be used at the next possible opportunity ( $\alpha$ persistence); at the same time, the more recently a generic non-separated verb-particle/preposition pattern (as in *I look at the house*) was used, the more likely the continuous variant becomes ( $\beta$ -persistence). Haddican and Johnson (2012) report a judgment experiment and a Twitter corpus study showing regional differences such that the split variant is more favored in British and Irish varieties of English than in North-American varieties. Paquot et al. (2019) analyze materials from the Louvain International Database of Spoken English Interlanguage to explore factors influencing particle placement choices for English as a Foreign Language (EFL) learners, compared to native speakers (as sampled in the Louvain Corpus of Native English Conversation). Conditional inference trees and conditional random forest analysis suggest that the probabilistic grammars of learners with Germanic L1s is similar to that of native speakers, while learners with non-Germanic L1s have significantly more simple grammars. Finally, Röthlisberger and Tagliamonte (2020) investigate particle placement in a corpus covering vernacular speech of six communities in Ontario. Phrasal verb tokens were annotated for two language-internal constraints and a number of language-external constraints. Regression analysis shows, among other things, that younger speakers use the continuous variant

more than older speakers, which is interpreted as evidence of a change in progress.

The analysis in this book is based on Grafmiller and Szmrecsanyi (2018) (see also Szmrecsanyi et al., 2016a; Tamaredo et al., 2019 for spin-off studies).

### Probabilistic Constraints on the Particle Placement Alternation

The particle placement alternation has been reported to be probabilistically constrained by a number of factors, the most important of which include the following:

- Consistent with the principle of end-weight (Behaghel, 1909; Wasow and Arnold, 2003), longer direct objects are reported to favor the continuous variant (Kennedy, 1920, 30; Quirk et al., 1985, 1154; Biber et al., 1999, 932-933).
- According to the literature, given language users' preference for placing old information before new information, if the direct object is discourse-given, the split variant is preferred. If it is discourse-new, there is a preference for the continuous variant (see e.g. Kruisinga and Erades, 1953; Chen, 1986).
- The literature suggests that concrete (and hence more accessible) direct objects favor the split variant (see e.g. Gries, 2003; Haddican and Johnson, 2012).
- We know that the split variant is preferred if the particle verb construction is followed by a directional prepositional phrase, as in send cattle off to the mainland (see e.g. Fraser, 1976; Gries, 2003).
- According to e.g. Gries (1999), Biber et al. (1999, 933), and Quirk et al. (1985, 1155), particle verb constructions with an idiomatic meaning (as in carry out duties) prefer the continuous variant, while constructions where the particle has literal/spatial meaning (as in carry garbage out) favor the split variant.

More information on how we operationalized the above factors is provided in Chapter 4.

#### 2.6 Summary

This book is concerned with "[v]ariability in the linguistic signal within a given language" (van Hout and Muysken, 2016, 250), also known as the availability of "alternate ways of saying 'the same' thing" (Labov, 1972, 188) in – crucially - the realm of grammar (that is, morphology and syntax). This chapter began with a discussion of how the existence of this type of grammatical variation has been and still is a controversial idea in various linguistic communities of practice. We pointed out that for us, this question is (and should be) an empirical

2.6 Summary 33

one, and the evidence shows that grammatical alternations are in fact plentiful in standard and nonstandard Englishes around the world. Against this backdrop, we reviewed the literature on the three grammatical alternations subject to study in this book: the genitive alternation, the dative alternation, and the particle placement alternation.