1

Introduction

Since the earliest work in generative grammar, theories of natural language syntax have proliferated and developed along disparate and often contradictory paths. Nevertheless, the content of the human language faculty, or at least the part of it pertaining to sentence-building, has continued to be the principal locus of investigation. Issues such as defining grammaticality, the 'poverty of the stimulus' problem, how to delimit the typology of possible languages, and the relation between so-called competence and performance remain central to the discussion, albeit with ostensibly little consensus across camps. One need look no further than the contrasting traditions of transformational approaches, which typically posit derived structures and movement as a fundamental syntactic mechanism (Chomsky 1965 and subsequent studies), and those which reject transformations in favour of enriched theories of the lexicon and feature-structural representations, such as Lexical-Functional Grammar (Bresnan 1982 and subsequent studies) and Head-Driven Phrase Structure Grammar (Pollard and Sag 1994). Yet a third dimension underlying the syntax debate is that of the connectionist versus computational or algebraic cognitive models, that is, whether the human language faculty is best viewed as a formal system of manipulating symbols, as a neural network which encodes weights learned from external stimuli, or as some fusion of the two (see Pater 2019 for a recent summary). Perhaps the most well-known example of an explicitly hybrid approach, adopting both formal representations and connectionist assumptions, is Optimality Theory (OT; Prince and Smolensky 1993), in which formal constraints are ranked

and violable, and the task of the learner is to acquire the language-specific ranking. Finally, a fundamental concern of syntactic theory has been the restrictiveness of Universal Grammar (UG), a topic which in the 'Principles and Parameters' tradition has centered on the problem of parameter-setting: how to account for the full typological range of attested languages and the acquisition thereof while minimising the formal machinery attributed to UG (see, e.g., Huang and Roberts 2016 and references therein).

This book is written in order to contribute to the above discussions from a unique perspective: one that combines both computationalist and connectionist assumptions, posits a rich feature-matching apparatus as opposed to transformations, integrates insights from the Minimalist literature (Chomsky 1995 and subsequent studies), and proposes a holistic framework for syntactic investigation. In particular, three strands of preceding literature are built upon here: an Optimality Theoretic approach to constraint interaction applied to syntax (see Grimshaw 2001, Legendre et al. 2001, among others), an approach to feature-mapping between grammatical levels based on Linking Theory (LT) (Kiparsky 1997, 2001), and a model of morphosyntactic variation and change (Kroch 1989a,b, 1994, Pintzuk 1999, among others). Although OT-based analyses are commonplace in phonology, there remains a relative dearth of in-depth studies in the syntax literature from an OT perspective. The second key aim of this volume is to provide substantial empirical support for the proposed theory via a detailed case study of morphosyntactic phenomena in two closely related languages, Icelandic and Faroese (see Þráinsson 2007, Þráinsson et al. 2004/2012 for recent treatments). A considerable amount of new data from the author's fieldwork conducted on the Faroe Islands and Iceland is presented with respect to the theoretical questions of focus. An additional purpose of this book is to provide a sufficiently explicit description of the core aspects of the theory to enable future researchers to test the framework against new data, including detailed discussion of methodological concerns. Significant efforts have been made to support the presentation of data with statistical rigour, including plots and figures where appropriate. The section on competing grammars also touches on machine learning approaches to modelling grammatical variation, attempting to bridge the perceived divide between neural network and algebraic formalisms while acknowledging the importance of sociolinguistic factors in the selection of variants.

While, in principle, any generative theory of syntax must speak to the full range of possible syntactic phenomena, special attention is given here to one particular subdomain, namely grammatical case. Since case touches on both the semantics–syntax and syntax–morphology interfaces, it represents a good testing ground for grammatical hypotheses. Zooming in yet further, the central problem explored in this book can be summarised by the following high-level question: how can we approach a unified account of case-marking, particularly in those instances where morphological case does not transparently map to grammatical function? The theory of case has been a topic of considerable debate and little consensus in the syntactic literature. The phenomenon of non-nominative subjects has been the focus of much discussion, particularly since the classic works on Icelandic (Andrews 1976, Levin and Simpson 1981, Zaenen et al. 1985 and subsequent studies). Indeed, the majority of theories of 'quirky' case have, to a greater or lesser extent, built upon a framework that assumes Icelandic as the archetype. However, a closely related language, Faroese, has been largely neglected despite it exhibiting case-marking patterns that differ from Icelandic in challenging ways. In the chapters which follow, it is argued that even apparently idiosyncratic Faroese and Icelandic case-marking patterns are in fact predictable from general principles, which in turn have implications for the kinds of case systems we expect to encounter cross-linguistically.

As noted above, a version of Kiparsky's LT is adopted here (Kiparsky 1997, 2001), which acknowledges both syntactic positions and case morphology as means of licensing arguments and does not collapse argument structure into syntax. Instead, information relevant to case assignment is encoded via features that link the levels of abstract, morphosyntactic and morphological case. Abstract case is defined by a hierarchy of theta-roles, which itself is derived from the Semantic Form of the verb (Bierwisch 1986, Wunderlich 1997). By assuming a separate abstract case representation distinct from syntactic positions, we allow for phenomena where a single grammatical function may be instantiated by several different morphological cases (e.g. subjects in Finnish) or indeed multiple positions (e.g. subjects in Icelandic). However, this additional generative capacity is not unconstrained: as highlighted in Section 2.1, syntactic configurations which license structural cases may share certain properties, such as being specifiers of a head in the extended verbal domain. These generalisations are

captured via interaction of constraints on phrase structure and feature identity, as laid out more fully in Chapter 9. By acknowledging the availability of positional licensing in some languages, we also capture the fact that mismatches between positional and inflectional case may be tolerated, as in Faroese and Icelandic. Moreover, by giving syntax access to inflection, that is, distinguishing morphological exponence from case within syntax, we cover instances where a syntactic operation appears to track morphology, such as the dependence of agreement on case. In the chapters which follow, I will show that the range of observed variation in argument realisation is both readily explained and appropriately constrained by the proposed model.

Also following Kiparsky (2001), matching between levels of case is implemented in OT (Prince and Smolensky 1993, McCarthy and Prince 1995), which provides a way of formalising the generalisation that many linguistic phenomena involve markedness hierarchies, that is, a default or 'elsewhere' form with potentially increasingly specific forms depending on the relevant grammatical conditions. Case is a prime example of this, where in accusative languages nominative is the unmarked subject case and non-nominative subjects the more marked form. Under OT, grammar is an optimisation of conflicting pressures, the set of which is universal, but the ranking of which is language-specific. This book explores the hypothesis that a set of appropriately ranked, violable constraints is able to account for a range of case-marking, agreement and word-order facts in Faroese and Icelandic and correctly generates case-marking patterns in other typologically disparate languages.

Moreover, this proposal integrates a competing grammars model of synchronic intra-linguistic variation that goes beyond mere descriptive adequacy, providing a framework for approaching morphosyntactic variables that incorporates both internal and external factors.[1] Morphosyntactic variation is attributed to a probabilistic calculus, in which grammars are selected from a set of rankings available to the speaker according to differently weighted factors that depend on the variable in question. It is possible to test the competing grammars hypothesis empirically by training a model on corpus data, learning the weights assigned to relevant factors and predicting when a speaker

---

[1] See Kroch (1989a,b, 1994), Santorini (1992, 1993), Pintzuk (1999), Wallenberg (2016), among others, for the origin of this idea in the context of language change and Fritzenschaft et al. (1990), Yang (2000, 2002), among others, for similar ideas in the acquisition literature.

is likely to select a given grammar. The accuracy of the model can be verified through logistic regression and basic machine learning techniques, an exciting avenue for future research. Therefore, this approach also presents an opportunity for developing computational methods to explore central questions in syntactic theory.

All of the above theoretical claims are buttressed by empirical evidence from extensive surveys conducted in the Faroe Islands and Iceland, as well as from corpora and native speaker consultations; the significance of the findings is demonstrated by repeatable statistical models of the patterns observed in the data. Thus, this volume aims to account not only for discrete variants, such as case selection or available argument positions in syntax, but also the kinds of grammatical, information-structural, sociolinguistic and contextual factors that contribute to case-marking in actual usage. The adoption of these three components – i.e. LT, OT and the Competing Grammars Model, hereafter abbreviated to *Optimal Linking Grammar (OLG)* – provides a cross-linguistically tractable framework for approaching case-marking phenomena that is not only descriptively adequate but is ultimately more explanatory than most contemporary approaches to case in the generative syntax literature.

1.1   The Puzzle of 'Quirky' Case

This section introduces a phenomenon which necessitates a detailed account of how grammatical relations, syntactic structure and morphology interact. Case-marking in Insular Scandinavian languages is an oft discussed topic, but there remain interesting questions to be answered, one of which is revealed by comparison of experiencer–stimulus predicates in the two languages. In both Faroese and Icelandic the standard transitive case-marking pattern is nominative–accusative, but some verbs occur with non-nominative higher arguments (1–2).

(1)   a.   Far.   *Eg   sá   gentan*
             I.NOM saw girl-the.ACC.SG
             'I saw the girl'

      b.   Far.   *Henni   manglar   mat*
             her.DAT lacks.3SG food.ACC.SG
             'She lacks food'

(2)   a.   Ice.   *Ég   sá   stelpan*
             I.NOM saw girl-the.ACC.SG
             'I saw the girl'

b. Ice. *Mér   ógna   þau   vindaský*
me.DAT terrify.3PL those.NOM.PL winds.NOM.PL
'I am terrified of those winds'

c. Ice. *Hana   vantar   peninga*
her.ACC lacks.3SG money.ACC.SG
'She lacks money'

The sets of verbs which mark subjects with non-nominative case overlap across the two languages, but the Icelandic set is much larger than the Faroese and with a greater variety of case frames (see Þráinsson 2007:156–172). Moreover, in Faroese verbs with accusative subjects are no longer commonly used, unlike in Icelandic (2c). The central empirical question addressed here is: why are Faroese and Icelandic sentences with dative subjects different with respect to their object case and agreement? The distinction is illustrated in (3).

(3)   a. Ice. *Mér   líka   hundar*
me.DAT like.3PL dogs.NOM.PL
'I like dogs'

b. Far. *Mær   dámar   hundar*
me.DAT likes.3SG dogs.ACC.PL
'I like dogs'

As can be seen in (3), in Icelandic the object argument in such sentences bears nominative case and triggers number agreement on the finite verb, whereas in Faroese the object bears accusative and occurs with default third person singular verb agreement.[2] On the surface, it is surprising that this phenomenon, a highly marked structure cross-linguistically, should exhibit such differences between two closely related languages, where the sentence type in question has the same origin in Old Norse sentences with preverbal datives (van der Gaaf 1904, Jespersen 1927, Allen 1995, Rögnvaldsson 1995, Barðdal and Eyþórsson 2003). Moreover, since it has long been known that Icelandic marks objects with nominative case in the presence of dative or genitive subject case, it is unexpected that Faroese marks the object with accusative in such predicates. Additionally, it remains to be explained why Icelandic sentences with non-nominative subjects

---

[2] It has also been observed that full person agreement is not possible with nominative objects in Icelandic (Sigurðsson 1991, 1996, Taraldsen 1995 and subsequent studies), but this fact is tangential to the case-marking difference, since the patterns can be explained solely in reference to number agreement; that object agreement is 'impoverished' relative to subject agreement is unsurprising given the markedness of object agreement more generally.

exhibit object agreement in number while the same apparent structure in Faroese exhibits non-agreement in number, or perhaps agreement with a null expletive (Barnes 1986, Þráinsson et al. 2004, 2012).

If the dative argument in (3) in both languages is a true subject by standard criteria, which does seem to be the case (Zaenen et al. 1985, Barnes 1986), and if it is corroborated by results from fieldwork presented in this volume, the difference in object case cannot rest upon a difference in subjecthood of the dative. It is interesting, however, that the difference in object case co-varies with a difference in number agreement. Therefore, the main hypothesis to investigate is that these facts are connected. The OLG account presented here posits that **the difference between Icelandic and Faroese dative-subject predicates results from a conflict between two pressures: (i) to mark the object with regular structural case, and (ii) to agree with an overt nominative argument**.[3] If these pressures are weighted differently in the two languages, with Icelandic preferring object agreement and Faroese preferring accusative structural case, the sentences in (3) have an explanation. Furthermore, such an account appeals to general principles rather than *ad hoc* idiosyncrasies and makes testable predictions about the typology of languages with case-marking.

In order to test this claim, two other reasonable hypotheses must first be ruled out:

(i) **Different structural object position**: if Icelandic and Faroese can be shown to have a distinct object position in these languages, *and* said position is shown to be associated with nominative case-marking in Icelandic, the difference could be attributed to the configuration of the object with respect to other clausal elements.

(ii) **Lexical case-marking**: if the Faroese accusative object case can be shown to be lexically assigned (i.e. associated with the subset of verb lexemes marking dative case on the subject), previous analyses of Icelandic could be retained, in which accusative case is unavailable due to some kind of 'nominative first' preference (e.g. Yip et al. 1987).

---

[3] By 'object' is meant the syntactic instantiation of [–HR] abstract case, which in these languages is standardly an argument which occupies object position (V,Comp). For the purposes of the constraints, an object is defined by the abstract case features, which different languages realise differently in morphosyntax.

In order to rule out these hypotheses, we must test whether the object in each language (a) behaves like a regular object with respect to its structural position and (b) bears structural or lexical case. One means of investigating (a) in Scandinavian languages is the phenomenon known as *object shift*: if the object in both languages behaves no differently with respect to object shift, this constitutes evidence for it being structurally the same as a standard transitive object. Regarding (b), it is possible to determine whether the case is structural or lexical by testing *case preservation* behaviour: in Icelandic, when an object marked with lexical case, such as dative, is passivised, the corresponding subject of the passive 'preserves' case and is not replaced by nominative. In contrast, structural object case (accusative in both languages) is replaced by nominative on the passive subject. If the Faroese verbs which mark accusative object case passivise and the subject of the passive is nominative, this is consistent with the case being structural and not lexical.

These phenomena were investigated in extensive fieldwork on the Faroe Islands and Iceland via surveys and consultations with native speakers of each language, the results of which suggest that (i) and (ii) are not viable explanations for the observed patterns. Moreover, the data collected *are* consistent with the OLG proposal, namely that the key difference is a preference in Icelandic for agreement with a nominative argument conflicting with a pressure to express structural object case, whose relative importance is reversed in Faroese. These results have implications beyond Scandinavian languages, since they indicate that similar conflicting pressures are responsible for case-marking and agreement patterns in multiple disparate language families. Indeed, it has already been shown that a very similar interaction of constraints can account for Indo-Aryan case-marking and changes in case and agreement systems within that family (Deo and Sharma 2006, Kiparsky 2017).

1.2   Theoretical Overview

The OLG proposal involves three central theoretical assumptions which are shown to be necessary to account for the range of data observed in Icelandic and Faroese alone and also provides a flexible enough framework to generate realistic typologies of case systems beyond Scandinavian. These three pillars build upon previous work in the

morphosyntax literature but also innovate in terms of the specifics of case theory and how grammar competition is modelled to capture intra-language variation.

1. **Linking Theory (LT)** (Kiparsky 1997, 2001): case is determined by semantics, syntax and morphology, and the *linking* between these levels determines the output.

2. **Optimality Theory (OT)** (Prince and Smolensky 1993): grammar is a *harmonic optimisation*, that is, a universal set of violable constraints with language-specific rankings.

3. **Competing grammars model (CGM)** (Kroch 1989a,b, 1994, Pintzuk 1999): native speakers have synchronic access to multiple competing grammars, where a grammar is defined by a constraint ranking; grammar selection is probabilistic.

This book posits that such a theoretical apparatus is in fact necessary for empirical reasons and makes better sense of the data than a theory which collapses all of case and agreement into the syntactic component. Moreover, it does not attribute the difference between Icelandic and Faroese to mere language-specific exceptions or idiosyncrasies but to general principles of language. Each component of the theory also makes testable predictions that are demonstrably borne out cross-linguistically.

1.2.1   Linking Theory

Originally proposed by Kiparsky (1997, 2001), the basic premise of LT is three distinct levels of case: abstract, morphosyntactic and morpho-logical. Abstract case is generated from a Semantic Form representation of the predicate and its argument structure. Morphosyntactic case is so called because languages may make use of *syntactic position*, *case inflection* or both to instantiate case within syntax. Finally, morpho-logical case is a representation of mapping morphosyntactic case to morphology, that is, the morphemic representation of case that feeds the pronounced surface form. All three of these levels are represented by the binary features [±H(ighest)R(ole)] and [±L(owest)R(ole)], which refer to a hierarchy of thematic roles. These features 'mean' something distinct at each level, since semantics, syntax and morphology manipu-late distinct types of elements: for example, [+HR] may be paraphrased

as 'most prominent argument' at abstract case, 'subject position' or 'nominative inflectional case' in morphosyntax, and 'nominative morpheme' at morphology. The theory also presupposes Lexicalism, in which word-formation is subject to pre-syntactic lexical constraints as opposed to syntactic transformations, and therefore words enter syntax fully inflected (Chomsky 1970, Halle 1973, Siegel 1974 and subsequent studies). Importantly, this does not rule out syntactic constraints targeting sub-parts of words, such as case or agreement morphemes; it simply rules out the *construction* of words by syntactic rules or constraints (e.g. some aspects of Distributed Morphology [DM], see Halle and Marantz 1993, 1994 and subsequent studies). In that sense, the theory adopted here joins a family of theories that combine lexicalism with a constraint-based architecture, such as Lexical–Functional Grammar (Bresnan 1982) or Head-Driven Phrase Structure Grammar (Pollard and Sag 1987, 1994). However, it should be noted that LT is not necessarily incompatible with many of the ideas espoused in DM-based approaches: for instance, since spellout is post-syntactic in DM, mismatches are also possible between morphological case and the features syntax operates on. In DM, such mismatches may also differ across languages according to differing inventories of pronounceable morphemes; thus, there is an optimised mapping between two domains (which in OLG is expressed in Optimality Theoretic terms). Hence, the LT component of OLG can be seen as a complementary proposal which builds upon ideas present in the literature rather than opposing all developments of DM. It is, however, opposed to some of the more derivation-based proposals and explicitly constrains the range of possible mismatches between levels.

Kiparsky (2001) provides evidence from Finnish for the necessity of three levels of case. The table in (4) shows the paradigms of structural cases for nouns and pronouns as typically presented in pedagogical grammars:

(4)

	Finnish structural cases 1			
	Nouns: 'bear'		Pronouns: 'you'	
	SG	PL	SG	PL
NOM	*karhu*	*karhu-t*	*sinä*	*te*
ACC	*karhu*, *karhu-n*	*karhu-t*	*sinu-t*	*te-i-dä-t*
GEN	*karhu-n*	*karhu-j-en*	*sinu-n*	*te-i-dä-n*
PART	*karhu-a*	*karhu-j-a*	*sinu-a*	*te-i-tä*

The distribution of the accusative singular in *–n* in the noun paradigm is formalised as Jahnsson's Rule, which can be paraphrased as 'verbs