

2. Strukturelle Komplexität in der Linguistik

2.1. Grundsätzliche Überlegungen zur linguistischen Komplexität

2.1.1. Wissenschaftsgeschichtlicher Überblick

Kusters (2003: 1–5) und Sampson (2009) geben einen ausgezeichneten Überblick darüber, ob, und wenn ja, auf welche Art und Weise, die Komplexität der Sprache in der Linguistik des 19. und 20. Jh. eine Rolle gespielt hat. Deswegen werden hier nur die wichtigsten Punkte kurz zusammengefasst, was natürlich eine gewisse Verallgemeinerung als Konsequenz hat. Die Übersicht ist chronologisch organisiert.

Im philosophischen Idealismus und der Romantik des 19. Jh. wurde das vermeintlich hohe Niveau europäischer Kultur mit der vorgeblich hohen Komplexität europäischer Sprachen in Verbindung gebracht (Kusters 2003: 2). Humboldt (1836) bringt dies wie folgt auf den Punkt: „Die Sprache ist gleichsam die äußerliche Erscheinung des Geistes der Völker; ihre Sprache ist ihr Geist und ihr Geist ihre Sprache [...]“ (Humboldt 1836: 53). Der angenommene Zusammenhang zwischen Sprache und Denken kann ebenfalls mit Humboldt illustriert werden, der davon ausging, dass flektierende Sprachen sich für komplexe Gedanken besser eignen als andere Sprachtypen (Kusters 2003: 3). Mit den Junggrammatikern gegen Ende des 19. Jh. hatten Denken und Kultur keine so große Bedeutung mehr in der Sprache, wie dies im Idealismus und in der Romantik der Fall war. Sprache wurde als Naturphänomen (vgl. Ausnahmslosigkeit der Lautgesetze) angesehen und die Verwandtschaften zwischen den Sprachen (vgl. Stammbaumtheorie) standen im Vordergrund (Kusters 2003: 3). Auch die Strukturalisten interessierten sich (zumindest in erster Linie) nicht für die Komplexität in der Sprache. Vielmehr ging es darum, ein umfassendes Instrument zu haben, um alle Sprachen beschreiben zu können (Kusters 2003: 4). In der generativen Grammatik spielte Komplexität ebenfalls keine Rolle, da Sprache als Teil der menschlichen Biologie und nicht der Kultur gesehen wird (Sampson 2009: 6). Deswegen steht die I-Language im Fokus der Betrachtung, die E-Language wird für die linguistische Analyse als eher unwichtig angesehen. Ziel ist es, Übereinstimmungen zwischen Sprachen und den Sprachen zugrunde liegenden Universalien herauszuarbeiten (Kusters 2003: 4). Unterschiede zwischen den Sprachen werden durch Unterschiede in der Derivation erklärt (z.B. Transformationen, Move und Merge etc.). Kürzlich erschienene Bände zeigen jedoch, dass linguistische Komplexität vermehrt auch innerhalb von generativen Modellen analysiert wird (z.B. Culicover 2013, Trotzke & Bayer 2015). Die Aufsätze in dem von Trotzke & Bayer (2015) herausgegebene Band beschäftigen sich vor allem mit syntaktischer Komplexität (z.B. Rekursion, Derivation, Einbettung etc.) und deren Erwerb, die Monografie von Culicover (2013) mit dem Sprachwandel, Spracherwerb und Prozessierung und deren Zusammenhang bzw. Einfluss auf linguistische Komplexität.

Im 20. Jh. stand also die Komplexität der Sprache nicht im Fokus des Interesses. Vielmehr galt, verallgemeinernd ausgedrückt, dass sich Sprachen in ihrer Komplexität nicht unterscheiden, was als *Equi-Complexity*-Hypothese bezeichnet wird. Dies kann auch als Gegenreaktion auf die Annahmen aus dem 19. Jh. verstanden werden. Gleichzeitig sprachen

besonders Variationslinguisten im 20. Jh. immer wieder von Komplexitätsunterschieden, z.B. Fergusons (1959) Unterscheidung zwischen *High* und *Low Varieties*. Darauf wird im anschließenden Kapitel noch genauer eingegangen.

Im 21. Jh. ist ein wachsendes Interesse an struktureller Komplexität in der Sprache zu beobachten. Dazu sind bereits etliche Aufsatzsammlungen und Dissertationen entstanden, wie, um nur einige Wenige zu nennen, Kusters (2003), Miestamo, Sinnemäki & Karlsson (2008), Sampson, Gil & Trudgill (2009), Sinnemäki (2011), Szmrecsanyi & Kortmann (2012). Dabei stehen vor allem folgende Fragen im Zentrum: Sind alle Sprachen gleich komplex? Wird höhere Komplexität in einem Subsystem (z.B. Morphologie) durch niedrigere Komplexität in einem anderen Subsystem (z.B. Syntax) ausgeglichen? Wie kann Komplexität gemessen werden? Wie können Unterschiede in der Komplexität zwischen Sprachen erklärt werden? (Baechler & Seiler 2016). Warum die linguistische Komplexitätsforschung seit wenigen Jahren und relativ plötzlich so viel Aufwind erfahren hat, liegt vermutlich in der klaren, expliziten Entkoppelung der früher angenommenen Verbindung zwischen ‘besonders komplex’ und ‘besonders wertvoll’, d.h., dass eine vor allem flexionsmorphologisch besonders komplexe Sprache auch besonders wertvoll sei. Versteht man jedoch eine Sprache als ein System von Regeln, die vorgeben, wie das Lexikon aufgebaut wird und wie mit diesem Lexikon Sätze generiert werden, müsste man – wenn schon – vom Gegenteil ausgehen. Denn ein solches System zur Informationsverarbeitung wäre wohl dann besonders elegant, wenn es mit sehr wenigen Regeln (also möglichst geringer Komplexität) auskommt, folglich sehr effizient ist.

2.1.2. Hypothese der gleichen Komplexität (*Equi-Complexity-Hypothese*)

Im Strukturalismus des 20. Jh. wurde generell angenommen – insofern sich die strukturalistischen Linguisten überhaupt dazu äußerten – dass sich Sprachen in ihrer Gesamtkomplexität nicht signifikant unterscheiden. Diese Hypothese wird *Equi-Complexity-Hypothese* genannt. Ihr prominentester Vertreter ist wohl **Hockett (1958)**, der feststellt, dass „[...] impressionistically it would seem that the total grammatical complexity of any language, counting both morphology and syntax, is about the same as that of any other. This is not surprising, since all languages have about equally complex jobs to do [...]“ (Hockett 1958: 180). Wenn man also die Komplexität aller Subsysteme messen würde, käme unter dem Strich heraus, dass alle Sprachen ungefähr gleich komplex sind. Hockett (1958) sagt jedoch nichts dazu, welche die relevanten Subsysteme sind und wie deren Komplexität gemessen werden kann. Sehr verwundert und pointiert äußert sich **Sampson (2009)** über den angenommenen Ausgleichsmechanismus in Sprachen:

If it really were so that languages varied greatly in the complexity of subsystem *X*, varied greatly in the complexity of subsystem *Y*, and so on, yet for all languages the totals from the separate subsystems added together could be shown to come out the same, then I would not agree with Hockett in finding this unsurprising. To me it would feel almost like magic. (Sampson 2009: 2–3)

Obwohl im 20. Jh. die *Equi-Complexity*-Hypothese dominierte, sprachen besonders Variationslinguisten immer wieder von Komplexitätsunterschieden zwischen Sprachen. Der einflussreichste Vertreter ist hier u.a. wohl Ferguson (1959), der *High Varieties* von *Low Varieties* differenziert: „One of the most striking differences between H[igh] and L[ow] in the defining languages is in the grammatical structure: H has grammatical categories not present in L and has an inflectional system of nouns and verbs which is much reduced or totally absent in L“ (Ferguson 1959: 333). Es bleibt aber offen, wie diese Unterschiede quantifiziert werden können. In der Folge sollen nun einige Studien vorgestellt werden, die zeigen, dass es Ausgleichstendenzen zwischen den Subsystemen innerhalb einer Sprache gibt, wie auch solche Studien, die das Gegenteil belegen. Diese Liste versteht sich als kurzen Überblick mit Fokus auf den Resultaten und erhebt keinen Anspruch auf Vollständigkeit.

Sowohl **Juola (2008)** als auch **Ehret & Szmrecsanyi (2016)** verwenden ein informationstheoretisches Maß (Kolmogorov-Komplexität und Ziv-Lempel-Komplexität), um Komplexität zu messen, womit die Menge an Informationen und Redundanzen ermittelt werden kann (Juola 2008: 93). Konkret wird Komplexität durch das Verzerren bestimmter Subsysteme und durch Komprimierung gemessen. Die Datengrundlage bilden große Textkorpora in sechs Sprachen: bei Juola (2008) die Bibel, bei Ehret & Szmrecsanyi (2016) das Markusevangelium. Juola (2008) kann zeigen, dass sich Sprachen in ihrer Komplexität nicht wesentlich unterscheiden (Juola 2008: 106), während die Resultate von Ehret & Szmrecsanyi (2016) mit einer vergleichbaren Methode das Gegenteil demonstrieren (Ehret & Szmrecsanyi 2016: 78). Beide Studien können aber mit genau derselben Messmethode einen Ausgleich zwischen morphologischer und syntaktischer Komplexität nachweisen (Juola 2008: 104; Ehret & Szmrecsanyi 2016: 79–80).

Sinnemäki (2008) untersucht in 50 Sprachen die Markierung von Agens und Patiens, wozu vier Strategien existieren, aber nur die letzten drei als strukturelle Strategien berücksichtigt werden: Lexikon, Wortstellung, *Head Marking* und *Dependent Marking* (Sinnemäki 2008: 68). Gemessen wird, wie oft eine bestimmte Kodierungsstrategie verwendet wird, um Agens und Patiens zu unterscheiden (Sinnemäki 2008: 72). Sinnemäki (2008) stellt fest, dass in einigen Subdomänen Ausgleichstendenzen zu beobachten sind, vor allem verallgemeinernd zwischen Wortstellung und *Dependent Marking* (Sinnemäki 2008: 84–85). Die meisten potentiellen Korrelationen jedoch, die geprüft wurden, waren äußerst gering oder inexistent, weshalb Trade-Offs als allgemeines Prinzip verworfen werden können (Sinnemäki 2008: 84).

Shosted (2006) überprüft in 32 Sprachen einen möglichen Ausgleich zwischen morphologischer und phonologischer Komplexität. Dazu zählt er die Anzahl möglicher Silben einer Sprache und die Anzahl der Marker in der Verbflexion (als eine Art Synthesegrad) (Shosted 2006: 9–17). Die Korrelation zwischen den Messresultaten ist leicht positiv, aber statistisch nicht signifikant (Shosted 2006: 1).

Auch **Nichols (2009)** konnte keine negativen Korrelationen zwischen verschiedenen Komponenten der Grammatik feststellen (Nichols 2009: 119). Sie untersucht 68 Sprachen und fünf Komponenten: Phonologie (u.a. Anzahl der Qualitätsunterschiede der Vokale), Synthese (u.a. Anzahl am Verb markierter Kategorien), Klassifikation (u.a. Genuskongruenz), Syntax (u.a. Anzahl unterschiedlicher Abfolgen von nominalen und pronominalen Argumenten sowie

dem Verb) und Lexikon (u.a. Anzahl Suppletivpaare in neun Paaren Vollverb/kausatives Verb) (Nichols 2009: 113).

Einen Ausgleich zwischen Subsystemen der Grammatik konnten nur die informationstheoretisch basierten Methoden nachweisen. Diese Methoden ermöglichen zwar im Prinzip gesamte Subsysteme zu messen. Sie sind jedoch probabilistisch und erlauben keinen Einblick in die Details der Ausgleichsmechanismen. Dies wäre jedoch wichtig, um die eventuellen Ausgleichsmuster zu verstehen und um ein Ausgleichsmuster abstrahieren zu können, denn so wäre dann ein Vergleich der Ausgleichsmuster von verschiedenen Sprachen möglich. Des Weiteren könnten diese Ausgleichsmuster und ihr Vergleich für die linguistische Theoriebildung nutzbar gemacht werden. In diese Richtung gehen eher Studien wie jene von Sinnemäki (u.a. 2008) und Nichols (u.a. 2009). Das Problem dieser Methoden jedoch ist, dass immer nur ein Ausschnitt einer Sprache oder eines Subsystems analysiert werden kann. Dafür ermöglichen sie aber einen detaillierteren Einblick in die inneren Vorgänge der Sprachen.

Dies tritt in Konflikt mit der *Equi-Complexity*-Hypothese, die implizit verlangt, dass die Gesamtkomplexität einer Sprache gemessen werden soll, d.h. alle Subsysteme. Erstens gibt es aber keinen Konsens über die Taxonomie der linguistischen Subsysteme. Zweitens stellt sich die Frage, wie die Komplexität der einzelnen Subsysteme gemessen werden kann, sodass eine Verrechnung der Resultate die Gesamtkomplexität darstellt. **Miestamo (2008)** bezeichnet dies als das Problem der Repräsentativität und als das Problem der Vergleichbarkeit:

The problem of representativity means that no metric can pay attention to all aspects of grammar that are relevant for measuring global complexity. Even if this were theoretically possible, it would be beyond the capacities of the mortal linguist to exhaustively count all grammatical details of the languages studied [...]. The problem of comparability is about the difficulty of comparing different aspects of grammar in a meaningful way, and especially about the impossibility of quantifying their contributions to global complexity. (Miestamo 2008: 30)

Beispielsweise stellt sich die Frage, wie viele Unterscheidungen im Aspektsystem gleich komplex sind und wie viele Unterscheidungen im Tempussystem (Miestamo 2006: 7). Daraus kann geschlossen werden, dass (zumindest vorerst) die Messung einzelner Subsysteme oder Teile von Subsystemen zu bevorzugen ist. Miestamo (2008) nennt dies *Local Complexity* im Gegensatz zur *Global* oder *Overall Complexity*, womit die Gesamtkomplexität einer Sprache gemeint ist (Miestamo 2008: 29). Viel grundsätzlicher geht es darum, dass nur Vergleichbares verglichen werden kann. Miestamo (2008) plädiert dafür, formale Aspekte der Grammatik (z.B. morphologische Systeme) von den funktionalen Aspekten (z.B. Kodierung von Tempus, Aspekt) zu unterscheiden und nur innerhalb dieser Bereiche einen Komplexitätsvergleich zwischen Sprachen vorzunehmen (Miestamo 2008: 31). Ein gutes Beispiel dafür ist die oben zitierte Arbeit von Sinnemäki (2008). Dabei wird die Markierung von Argumenten gemessen, wofür die Sprache unterschiedliche Mittel (z.B. morphologische und syntaktische) zu Verfügung hat.

2.1.3. Hat Komplexität eine Funktion in der Sprache?

Die *Equi-Complexity*-Hypothese, die vorwiegend auf den Strukturalismus zurückgeht, nimmt an, dass alle Sprachen gleich komplex sind und höhere Komplexität in einem Subsystem durch niedrigere Komplexität in einem anderen Subsystem kompensiert wird. Dies impliziert, dass Komplexität eine Funktion innerhalb der Sprache erfüllt. Im Kapitel 2.1.1. wurde gezeigt, dass besonders in der ersten Hälfte des 19. Jh. davon ausgegangen wurde, dass die strukturelle Komplexität einer Sprache mit der (u.a. kulturellen) Entwicklung ihrer Sprecher korreliert. Zusammenfassend kann man also Folgendes festhalten, was natürlich eine Verallgemeinerung darstellt. Sowohl im 19. Jh. als auch im 20. Jh. wurde angenommen, dass strukturelle Komplexität eine Funktion in der Sprache innehat. Im 19. Jh. wurde dies verwendet, um die angebliche Fortschrittlichkeit besonders Europas zu erklären. Im 20. Jh. ging man davon aus, dass die kommunikativen Anforderungen an eine Sprache immer dieselben sind, folglich alle Sprachen den gleichen Grad an Komplexität aufweisen müssen. Ein Beispiel dafür ist die bereits oben zitierte Aussage von Hockett (1958), der behauptet, dass Ausgleichstendenzen in der Komplexität von verschiedenen Subsystemen zu erwarten sind, „[...] since all languages have about equally complex jobs to do [...]“ (Hockett 1958: 180). Dem entgegnet Sampson (2009) mit einer ganz grundsätzlichen Überlegung: „[...] but it seems to me very difficult to define the job which grammar does in a way that is specific enough to imply any particular prediction about grammatical complexity“ (Sampson 2009: 2).

Es ist also sicher nicht möglich, jene Aufgaben umfassend zu beschreiben, die eine Sprache erfüllen muss. Noch grundlegender ist jedoch die Frage, ob eine komplexe Aufgabe zwangsläufig mit einer komplexen Grammatik gelöst werden muss: Weshalb sollten komplexe Konzepte auch nur durch eine komplexe Grammatik ausgedrückt werden können und umgekehrt, einfache Konzepte durch eine einfache Grammatik? Können wir doch mit derselben Sprache sowohl einfache wie auch schwierige Dinge auf eine mehr oder weniger komplexe Art formulieren. Folglich gibt es zwischen der Komplexität der Grammatik und der Komplexität der Dinge, die wir tun, keinen Zusammenhang. Dass Grammatik oft eben gerade nicht funktional ist, bringt Gil (2009) wie folgt auf den Punkt:

These facts cast doubt on a central tenet of most functionalist approaches to language, in accordance with which grammatical complexity is there to enable us to communicate the messages we need to get across. In spite of overwhelming evidence showing that diachronic change can be functionally motivated, the fact remains that language is hugely dysfunctional. Just think of all the things that it would be wonderful to be able to say but for which no language comes remotely near to providing the necessary expressive tools. For example, it would be very useful to be able to describe the face of a strange person in such a way that the hearer would be able to pick out that person in a crowd or a police line-up. But language is completely helpless for this task, as evidenced by the various stratagems police have developed, involving skilled artists or, more recently, graphic computer programs, to elicit identifying facial information from witnesses – in this case a picture actually being worth much more than the proverbial thousand words. Yet paradoxically, alongside all the things we'd like to say but can't, language also continually forces us to say things that we don't want to say; this happens whenever an obligatorily marked grammatical category leads us to specify something we would rather leave unspecified. English, famously, forces third person singular human pronouns to be either masculine or feminine; but in many contexts we either don't know the person's gender or actually wish to leave it unspecified [...]. (Gil 2009: 32)

Strukturelle Komplexität hat folglich nichts mit der Effizienz und Expressivität einer Sprache als Mittel zur Kommunikation zu tun (Miestamo 2006: 2). Auch ist sie kein Symptom für Zivilisation und Fortschrittlichkeit. Vielmehr kann Grammatik als ein System gesehen werden, dessen Variation in der Komplexität primär systeminterne Ursachen hat:

Rather than having evolved in order to enable us to survive, sail boats, and do all the other things that modern humans do, most contemporary grammatical complexity is more appropriately viewed as the outcome of natural processes of self-organization whose motivation is largely or entirely system-internal. In this respect, grammatical complexity may be no different from complexity in other domains, such as anthropological complexity, economics, biology, chemistry, and cosmology, which have been suggested to be governed by general laws of nature pertaining to the evolution of complex systems. (Gil 2009: 32–33)

Wenn Grammatik also ein sich selbst organisierendes System ist, dann besteht ebenfalls die Möglichkeit, dass auch die Komplexität der Grammatik verschiedener Sprachen unterschiedlich hoch ist. Die grammatischen Systeme von Sprachen mögen zwar untereinander mehr Ähnlichkeiten aufweisen als im Vergleich mit z.B. chemischen Systemen. Trotzdem gibt es keinen Grund, weshalb die grammatischen Systeme aller Sprachen gleich komplex sein sollten.

2.2. Strukturelle Komplexität: Definitionen, Messmethoden und Einflussfaktoren

Dieses Kapitel gibt einen Überblick über die Arbeiten, die Unterschiede in der strukturellen Komplexität u.a. durch sprachexterne Faktoren zu erklären versuchen. Alle bereits publizierten Untersuchungen zu diskutieren, würde den hiesigen Rahmen sprengen. Vielmehr soll ein Ausschnitt jener Studien vorgestellt werden, die sich mit den sprachexternen Faktoren beschäftigen, welche für die vorliegende Arbeit zentral sind: kleine, isolierte Sprachgemeinschaften mit wenig Sprachkontakt und einem engen Netzwerk vs. große, nicht isolierte Sprachgemeinschaften mit viel Sprachkontakt (vielen L2-Lernern) und losen Netzwerken. Unterschiedliche Arbeiten dazu werden in den Kapiteln 2.2.3. und 2.2.4. vorgestellt. Zuerst sollen jedoch die ersten publizierten Überlegungen zu möglichen Einflussfaktoren kurz eingeführt werden (Kap. 2.2.1.). Anschließend werden Arbeiten erörtert, in denen sprachexterne Faktoren herangezogen werden, die in der vorliegenden Arbeit nicht im Fokus stehen, die aber die Diskussion über den Zusammenhang zwischen struktureller Komplexität und sprachexternen Faktoren beeinflusst haben (Kap. 2.2.2.). Es handelt sich dabei um die Faktoren Alter, Geschlecht, Schicht, Region und Bevölkerungsgröße. In diesem Überblick stehen folgende Fragen im Vordergrund: Was wird unter struktureller Komplexität verstanden und wie kann diese gemessen werden? Welche Sprachen/Varietäten und welche linguistischen Beschreibungsebenen werden untersucht? Wie werden die Unterschiede in der Komplexität von Sprachen erklärt? In einem abschließenden Kapitel wird zuerst die mittlerweile etablierte Unterscheidung zwischen absoluter und relativer Komplexität vorgestellt. Anschließend wird eine detailliertere Unterscheidung von Komplexitätstypen eingeführt, die auf den Philosophen Rescher (1998) zurückgeht und von

Miestamo, Sinnemäki & Karlsson (2008) für linguistische Phänomene adaptiert wurde. Zuletzt werden die in den vorangehenden Kapiteln 2.2.1.–2.2.4. verwendeten Definitionen struktureller Komplexität zusammengefasst und kategorisiert (Kap. 2.2.5.).

2.2.1. Erste Überlegungen zu möglichen Einflussfaktoren

Überlegungen zu möglichen sprachexternen Faktoren, die mit der Variation in der strukturellen Komplexität zusammenhängen könnten, sind schon relativ alt. Einen Überblick über die ersten Überlegungen in diesem Zusammenhang bieten Baechler & Seiler (2016), woran ich mich in der Folge orientiere.

Der vielleicht erste Linguist, der soziale und geografische Faktoren mit der Höhe struktureller Komplexität in Verbindung brachte, ist Roman **Jakobson (1929)**. Er untersuchte das phonologische System ukrainischer Dialekte und stellte Folgendes fest: Zentrale, sich ausbreitende Dialekte, die von einer homogenen Sprachgemeinschaft gesprochen werden, haben ein kleineres Vokalinventar als die Dialekte an der Peripherie des ukrainischen Sprachgebiets (Jakobson 1929: 73).

Cette différence est due, en premier lieu à la tendance conservatrice qui est caractéristique des parlers de la périphérie, et en second lieu à des différences fonctionnelles. Il n'est pas rare d'observer que la tendance à simplifier le système phonologique croît à mesure que grandit le rayon d'emploi d'un dialecte, avec la plus grande hétérogénéité des sujets parlant la langue généralisée. On n'a pas encore, en linguistique, prêté assez attention à la différence essentielle de structure et d'évolution qui existe entre les parlers gravitant vers le rôle de *κοινή* ou langue commune, et ceux d'usage purement local. (Jakobson 1929: 73)

[Dieser Unterschied ist primär bedingt durch die konservierende Tendenz, die charakteristisch für die Sprachen an der Peripherie ist, aber auch durch funktionelle Unterschiede. Es ist nicht selten zu beobachten, dass die Tendenz, das phonologische System zu simplifizieren, zunimmt, je mehr das Areal, in dem der Dialekt gesprochen wird, größer wird, und mit einer heterogenen Sprachgemeinschaft, die die *Lingua franca* spricht. In der Linguistik hat man dem wesentlichen Unterschied in der Struktur und der Evolution noch nicht genügend Aufmerksamkeit geschenkt, der zwischen den Sprachen existiert, die in Richtung *Koiné* oder Gemeinsprache tendieren, und jenen, die ausschließlich lokal verwendet werden.] [meine Übersetzung]

Jakobson (1929) beschreibt hier also geografische und soziale Faktoren, die auf die Funktion einer Sprache Einfluss haben, was wiederum auf den Sprachwandel wirkt. **Hymnes' (1975)** Beobachtungen weisen in eine ähnliche Richtung wie jene von Jakobson (1929), wobei bei Hymnes (1975) soziale Faktoren der Sprachgemeinschaft im Vordergrund stehen. Er stellt fest, dass kleine Sprachen mit einem engen Netzwerk eine höhere strukturelle Komplexität aufweisen: „This latter process may have something to do with the fact that the surface structures of languages spoken in small, cheek-by-jowl communities so often are markedly complex, and the surface structures of languages spoken over wide ranges less so“ (Hymnes 1975: 50).

Auch **Werner (1975)** beobachtet Ähnliches wie Jakobson (1929). Er vergleicht kleine und isolierte Sprachgemeinschaften, die Jakobsons peripheren Sprachen entsprechen, mit Sprachen mit viel Kontakt, die Jakobsons sich ausbreitenden Sprachen gleichkommen. Bezüglich der möglichen Korrelation zwischen diesen Sprachgemeinschaftstypen und der strukturellen Komplexität einer Sprache stellt Werner (1975) dasselbe wie Jakobson (1929) fest: „Es ist – vermute ich – ganz allgemein ein Kennzeichen kleinerer, isolierter Sprachgemeinschaften, dass sie lange komplizierte Regelsysteme bewahren; großräumige Sprachkontakte und damit verbundene Interferenzen fordern dagegen die Analogien, wie sie ja auch von Kindern und Ausländern gerne gemacht werden“ (Werner 1975: 791).

Eine der ersten Arbeiten, in der explizit ein möglicher Zusammenhang zwischen sprachexternen Faktoren und struktureller Komplexität systematischer geprüft wird, ist **Braunmüllers 1984** erschienener Aufsatz. Er untersucht die Flexionsmorphologie des Isländischen, Färöischen und Friesischen. Dabei handelt es sich um Sprachen, die nicht nur wenig Sprachkontakt aufweisen (mit Nicht-Muttersprachlern wird in einer anderen Sprache kommuniziert), sondern auch geografisch isoliert sind: Alle drei Sprachen werden auf Inseln gesprochen, Friesisch zusätzlich an der Küste Schleswig-Holsteins. Braunmüller (1984) zeigt, dass viele unterschiedliche phonologische Regeln gleichzeitig wirken und so Opazität im Paradigma verursachen. Die entstandene Opazität im Paradigma wird nur sehr wenig durch Analogien ausgeglichen, was besonders bei kleinen und isolierten Sprachen vorkommt (Braunmüller 1984: 49). Dies wird vor allem durch zwei Charakteristika dieser Sprachgemeinschaften begünstigt. Erstens findet der Kontakt mit Nicht-Muttersprachlern in einer anderen Sprache statt, sodass diese Sprachen nur sehr selten als L2 erlernt werden und sie nicht als Koiné dienen (Braunmüller 1984: 49). Dies hat zur Konsequenz, dass Sprachen von kleinen und isolierten Sprachgemeinschaften nur wenig durch andere Sprachen beeinflusst werden und kaum L2-Simplifizierungen aufweisen. Zweitens sind kleine und isolierte Sprachgemeinschaften sehr homogen und deswegen die Sprache wenig strenger Normierung ausgesetzt, die in den Sprachwandel eingreifen könnte (Braunmüller 1984: 49). Braunmüller (1984) bringt also frühere Beobachtungen zu möglichen Einflussfaktoren zusammen: kleine Sprachgemeinschaften mit einem engen Netzwerk und hoher Homogenität, wenig L2-Lerner, kleinräumiger Gebrauch der Sprache, geografisch isoliert und peripher.

2.2.2. Sprachexterne Faktoren

In diesem Kapitel sollen drei Studien vorgestellt werden, die den Zusammenhang zwischen sprachexternen Faktoren und struktureller Komplexität untersuchen. Bei Sampson (2001) handelt es sich bei den sprachexternen Faktoren um Alter, Geschlecht, Schicht und Region, bei Hay & Bauer (2007) sowie bei Sinnemäki (2009) um Bevölkerungsgrößen.

Sampson (2009) prüft den hypothetischen Zusammenhang zwischen syntaktischer Komplexität und Alter, Geschlecht, Schicht sowie Region, wobei es sich also um demografische Daten handelt. Die Datengrundlage bildet sein eigens erstelltes Korpus ‘Christine’, das spontansprachliche Äußerungen enthält und auf dem British National Corpus basiert (Sampson 2009: 57–58). Syntaktische Komplexität wird als die Einbettungstiefe

definiert: „For a sentence to be ‘simple’ or ‘complex’ in traditional grammatical parlance refers to whether or not it contains subordinate clause(s)“ (Sampson 2009: 58). Da in der Spontansprache der Anfang und das Ende von Sätzen nicht immer eindeutig bestimmt werden können, wird die Einbettungstiefe auf der Ebene des Wortes definiert: „The present research treats degree of embedding as a property of individual words. Each word is given a score representing the number of nodes in the CHRISTINE ‘lineage’ of that word [...] which are labeled with clause categories“ (Sampson 2009: 59). In einem ersten Durchgang fließen die Äußerungen aller Informanten in die Analyse ein, in einem zweiten Durchgang nur die Äußerungen jener Informanten, die älter als 16 Jahre sind. Zwischen den fünf untersuchten Regionen (Süd- und Nordengland, Wales, Schottland, Nordirland) gibt es keinen signifikanten Unterschied in der Komplexität, unabhängig davon, ob die unter 16-Jährigen mit eingeschlossen sind oder nicht (Sampson 2009: 62, 64, 66). Dasselbe gilt bezüglich der sozialen Schicht (Sampson 2009: 64); die Resultate werden aber annähernd signifikant, wenn die unter 16-Jährigen ausgeschlossen werden (von mehr zu weniger komplex): ausgebildet-handwerklich, leitende und technische Berufe, ausgebildet-nicht handwerklich, teils ausgebildet/ohne Ausbildung (Sampson 2009: 67). Dieses doch eher erstaunliche Ergebnis die soziale Schicht betreffend wird dadurch erklärt, dass die Angaben zur Schicht in diesen Daten am wenigsten verlässlich sind (Sampson 2009: 67). Frauen produzieren etwas häufiger komplexe Sätze (signifikant) als Männer (Sampson 2009: 65). Werden jedoch nur die über 16-Jährigen berücksichtigt, ist das Resultat nicht mehr signifikant (Sampson 2009: 66). Das signifikante Ergebnis wird als Problem der Datengrundlage gedeutet, denn in diesem Sample sind mehr Männer als Frauen (Sampson 2009: 66–67). Die verschiedenen Altersgruppen unterscheiden sich signifikant in der syntaktischen Komplexität, wobei mehr komplexe Äußerungen produziert werden, je älter die Informanten sind (Sampson 2009: 65). Von der Annahme ausgehend, dass sich die Informanten bis 13 Jahre noch im Spracherwerbsprozess befinden, werden in einem weiteren Test die unter 13-Jährigen ausgeschlossen (Sampson 2009: 67–70). Auch in diesem Test zeigt sich, dass die syntaktische Komplexität mit dem Alter zunimmt (Sampson 2009: 70). Sampson (2009) interpretiert, „[...] that (while the evidence is not overwhelming) increase in average grammatical complexity of speech appears to be a phenomenon that does *not* terminate at puberty, but continues throughout life“ (Sampson 2009: 70).

Sinnemäki (2009) misst die Komplexität in der Markierung von Agens und Patiens in 50 Sprachen und prüft den Zusammenhang zwischen Komplexität und Bevölkerungsgröße. Zur Markierung von Agens und Patiens stehen drei Strategien zur Verfügung: *Head Marking* (am Verb), *Dependent Marking* (an Agens oder Patiens) und die Wortstellung (Sinnemäki 2009: 130). Komplexität wird als die Verletzung des Eine-Bedeutung-Eine-Form-Prinzips definiert. Von diesem Prinzip gibt es zwei Abweichungen, welche die Komplexität erhöhen: Verletzung der Ökonomie und Verletzung der Distinktheit (Sinnemäki 2009: 132). Wird das Prinzip der Ökonomie verletzt, wird zu viel markiert, d.h., die Sprache verwendet mehr als eine Strategie, um Agens und Patiens zu markieren, welche nicht komplementär verteilt sind (Sinnemäki 2009: 133). Wird das Prinzip der Distinktheit verletzt, wird zu wenig markiert, d.h., die Sprachen „[...] use only one strategy but in limited contexts, or allow a lot of syncretism in the head or dependent marking paradigms“ (Sinnemäki 2009: 133). Sinnemäki (2009) listet und zählt die Sprachen, die dem Eine-Bedeutung-Eine-Form-Prinzip entsprechen, und jene

Sprachen, die davon abweichen. Die Resultate können wie folgt zusammengefasst werden. Erstens nimmt die Anzahl der Sprachen, die dem Eine-Bedeutung-Eine-Form-Prinzip entsprechen, parallel zur Größe der Sprachgemeinschaft zu (Sinnemäki 2009: 135). Zweitens tendieren Sprachen, die von weniger als 10.000 Sprechern gesprochen werden, dazu, entweder Ökonomie oder Distinktheit zu verletzen. Dabei nimmt die Anzahl der Sprachen, die Ökonomie oder Distinktheit verletzen, ab, wenn die Anzahl der Sprecher zunimmt (Sinnemäki 2009: 135). Da es problematisch ist zu bestimmen, was eine große oder kleine Sprachgemeinschaft ist, wird mit demselben Sample ein weiterer Test gemacht. Dazu werden Grenzwerte bestimmt: Angefangen wird bei 250 Sprechern, der nächste Grenzwert ist das Doppelte des vorangehenden Grenzwertes usw. (Sinnemäki 2009: 135). Es kann festgestellt werden, dass Sprachen, die von einer Sprachgemeinschaft gesprochen werden, die kleiner oder gleich groß wie ein Grenzwert ist, eher Distinktheit oder Ökonomie verletzen. Des Weiteren tendieren Sprachen, die von einer Sprachgemeinschaft gesprochen werden, die größer als ein Grenzwert ist, dem Eine-Bedeutung-Eine-Form-Prinzip zu entsprechen (Sinnemäki 2009: 136–138). Zusammengefasst kann also festgehalten werden, dass Sprachen einer kleineren Sprachgemeinschaft in der Markierung von Agens und Patiens höhere Komplexität aufweisen als Sprachen einer größeren Sprachgemeinschaft.

Hay & Bauer (2007) prüfen einen Zusammenhang zwischen der Größe des Phoneminventars einer Sprache und der Größe der Sprachgemeinschaft. Das Sample beträgt 216 Sprachen (Hay & Bauer 2007: 388). Die Größe des Phoneminventars wird durch die Anzahl folgender Phonemkategorien gemessen: Basismonophthonge (unterscheiden sich in ihrer Qualität), Extramonophthonge (Unterscheidung der Länge und Nasalisierung), Diphthonge, Obstruenten und Sonoranten (Hay & Bauer 2007: 389). Getestet wird ein Zusammenhang zwischen diesen einzelnen Kategorien und der Bevölkerungsgröße wie auch zwischen der Gesamtgröße des Phoneminventars und der Bevölkerungsgröße. Es wird eine positive Korrelation sowohl zwischen jeder Phonemkategorie und der Bevölkerungsgröße als auch zwischen der Gesamtgröße des Phoneminventars und der Bevölkerungsgröße festgestellt (Hay & Bauer 2007: 389–390). Wichtig ist hier auch die Beobachtung, dass das Vokalinventar und das Konsonanteninventar jedoch miteinander nicht korrelieren (Hay & Bauer 2007: 391). Des Weiteren handelt es sich hier um eine statistische Tendenz. Es gibt also auch Sprachen, die gegen diese Tendenz laufen, wie z.B. das Färöische, das von relativ wenigen Sprechern gesprochen wird, aber ein großes Phoneminventar aufweist (Hay & Bauer 2007: 390). In einem weiteren Test wird untersucht, ob diese Tendenz verursacht wird von einem Zusammenhang zwischen Sprachfamilien und der Bevölkerungsgröße (Hay & Bauer 2007: 391). Tatsächlich hat die Sprachfamilie einen Einfluss, z.B. haben die indo-germanischen Sprachen die größten Phoneminventare, während die austronesischen Sprachen die kleinsten Phoneminventare aufweisen (Hay & Bauer 2007: 392). Die Bevölkerungsgröße ist jedoch neben der Sprachfamilie ein zusätzlicher separater, signifikanter Prädiktor für die Größe des Phoneminventars (Hay & Bauer 2007: 392). Weshalb Sprachen mit kleinen Sprachgemeinschaften auch kleine Phoneminventare aufweisen, kann nicht definitiv erklärt werden. Hay & Bauer (2007) diskutieren verschiedene Erklärungsversuche, auf die hier nicht weiter eingegangen werden kann. Eine Hypothese soll hier aber kurz vorgestellt werden. Trudgill (2004a) geht davon aus, dass kleine Sprachgemeinschaften entweder große oder kleine Phoneminventare aufweisen, während große Sprachgemeinschaften eher mittlere

Phoneminventare favorisieren (Trudgill 2004a: 317, zitiert aus Hay & Bauer 2007: 396). Große Phoneminventare werden dadurch erklärt, dass kleine Sprachgemeinschaften die Fähigkeit haben „[...] to encourage continued adherence to norms from one generation to another, however complex they may be“ (Trudgill 2004a: 317, zitiert aus Hay & Bauer 2007: 396). Kleine Phoneminventare dagegen werden auf das gemeinsame Wissen kleiner Sprachgemeinschaften zurückgeführt: „[...] initial small community size [...] would have led in turn to tight social networks, which would have implied large amounts of shared background information – a situation in which communication with relatively low level of phonological redundancy would have been relatively tolerable“ (Trudgill 2002: 720, zitiert aus Hay & Bauer 2007: 396). Den Zusammenhang zwischen phonologischer Komplexität, d.h. kleinem Inventar, und kleiner Sprachgemeinschaft illustriert Trudgill (2011) anhand des Hawaiischen (basierend auf Maddieson (1984)), das fünf Vokale, acht Konsonanten, eine CVCV-Silbenstruktur und nur 162 mögliche Silben hat:

My suggestion is that possessing only a small number of available syllables – and therefore a relatively small amount of redundancy – may, other things being equal, lead to greater communicative and/or cognitive difficulty because of a lack of contrastive possibilities. I suggest that while this lack of contrastive possibilities is entirely unproblematical for native speakers, languages such as Hawai’ian will cause difficulties for non-natives. Languages with very small phoneme inventories cause problems of *memory load* for foreign learners – they are L2 difficult. [...] The problem lies in the relative lack of distinctiveness between one vocabulary item and another, due to the necessarily high proportion of usage of possible syllables [...]. (Trudgill 2011: 124)

2.2.3. Soziale Isolation vs. Sprachkontakt

In diesem Kapitel werden einige Studien vorgestellt, die strukturelle Komplexität in Sprachgemeinschaften mit viel und wenig Sprachkontakt untersuchen. Dabei werden Sprachgemeinschaften mit wenig Sprachkontakt als isolierte Sprachgemeinschaften bezeichnet. Es handelt sich hier nur um einen exemplarischen Ausschnitt, in dem die wichtigsten Richtungen aufgezeigt werden sollen.

Peter Trudgill hat eine Vielzahl an Arbeiten veröffentlicht, die sich mit dem Zusammenhang zwischen struktureller Komplexität und dem Typ Sprachgemeinschaft beschäftigen. In Baechler & Seiler (2016) wurden die wichtigsten Resultate zusammengefasst, was hier in leicht abgeänderter Form übernommen wird. **Trudgill** unterscheidet vor allem zwei Typen an Sprachgemeinschaften. Der eine Typ wird wie bei Braunmüller (1984) durch geringe Bevölkerungsgröße, geografische und vor allem soziale Isolation charakterisiert (Trudgill 1992). Mit sozialer Isolation sind wenig Sprachkontakt und wenige L2-Lerner gemeint. Zusätzlich berücksichtigt Trudgill (2011) aber auch interne Charakteristika einer Sprachgemeinschaft: enge soziale Netzwerke, hohe soziale Stabilität und eine große Menge gemeinsamen Wissens (Trudgill 2011: 146). Folglich gibt es auf der einen Seite kleine, geografisch isolierte, stabile Sprachgemeinschaften mit einem engen Netzwerk und wenig Sprachkontakt, welche hier isolierte Sprachgemeinschaften genannt werden, und auf der anderen Seite große, geografisch nicht isolierte, sich verändernde Sprachgemeinschaften mit

losen Netzwerken und viel Sprachkontakt, welche hier als nicht isoliert gekennzeichnet werden. Diese zwei Typen an Sprachgemeinschaften bilden natürlich die beiden Pole, die Ausprägung der Charakteristika (z.B. hohe vs. geringe Bevölkerungszahl) wie auch die Charakteristika selbst können in unterschiedlicher Kombination auftreten (Trudgill 2011: 147). Einfachheitshalber wird hier jedoch weiterhin von den beiden Extremen, d.h. von isolierten und nicht isolierten Sprachgemeinschaften gesprochen. Zusammengefasst geht Trudgill davon aus, dass isolierte Sprachgemeinschaften dazu tendieren, höhere strukturelle Komplexität aufzuweisen und nicht isolierte Sprachgemeinschaften dazu, geringere strukturelle Komplexität zu zeigen. Die Gründe dafür sieht Trudgill im Sprachwandel, dessen Mechanismen und Ergebnisse abhängig vom Typ Sprachgemeinschaft sind, was hier kurz ausgeführt werden soll. Trudgill nimmt an, dass isolierte Sprachgemeinschaften höhere strukturelle Komplexität haben, weil in diesen Sprachgemeinschaften die Wahrscheinlichkeit groß ist, „[...] to find not only the preservation of complexity but also an increase in complexity, i.e. irregularity, opacity, syntagmatic redundancy, and non-borrowed morphological categories“ (Trudgill 2011: 64). Sowohl der Erhalt von Komplexität als auch die Zunahme an Komplexität können durch Charakteristika des Sprachwandels erklärt werden, der durch die Struktur der Sprachgemeinschaft beeinflusst ist. Der Erhalt von Komplexität liegt besonders an drei Faktoren. Erstens ist es einfacher für kleine Sprachgemeinschaften mit wenig Sprachkontakt und engen sozialen Netzwerken „[...] to enforce and reinforce the learning and use of irregularities“ (Trudgill 1992: 204) und deshalb „[...] to enforce and reinforce the learning and use of complexities by children and adolescents“ (Trudgill 1996: 13). Zweitens unterscheidet sich das Tempo des Sprachwandels in Abhängigkeit vom Typ der Sprachgemeinschaft: „In small, isolated, stable communities, linguistic change will be slower“ (Trudgill 2011: 103). Drittens sind kleine und isolierte Sprachgemeinschaften weniger vom Sprachwandel betroffen bzw. die „[g]eographically peripheral varieties which have been least subject to dialect contact most strongly resist [...] language change leading to simplification“ (Trudgill 1996: 6). Wenn jedoch Sprachwandel stattfindet, so sind kleine, isolierte Sprachgemeinschaften mit engen Netzwerken in höherem Maße fähig einerseits, „[...] to push through, enforce and sustain changes of a less natural or usual phonological type [...]“ (Trudgill 1996: 11), und andererseits „[...] [to] promote the spontaneous growth of morphological categories [...]“ (Trudgill 2009: 109). Allgemein kann folglich davon ausgegangen werden, dass in diesen Sprachgemeinschaften der Sprachwandel eher Komplexifizierung verursacht (Trudgill 2011: 103). Trudgill (2011) nennt diesen Typ der Komplexifizierung „[...] spontaneous, non-additive complexification“ (Trudgill 2011: 71). Es kann also festgehalten werden, dass kleine, geografisch isolierte, stabile Sprachgemeinschaften mit einem engen Netzwerk und wenig Sprachkontakt dazu tendieren, Komplexität zu erhalten wie auch zu erhöhen. Im Gegensatz dazu tendieren große, geografisch nicht isolierte, sich verändernde Sprachgemeinschaften mit losen Netzwerken und viel Sprachkontakt dazu, ihre Grammatik zu vereinfachen, „[...] because high irregularity, low transparency, and high levels of redundancy make for difficulties of learning and remembering for adolescent and adult learner-speakers“ (Trudgill 2009: 101). Es kann aber noch eine weitere Art des Kontakts beobachtet werden, nämlich „[...] long-term co-territorial contact situations involving child bilingualism“ (Trudgill 2011: 34). In dieser Kontaktsituation kommt Komplexifizierung in der Form von *Additive Borrowings* vor (Trudgill 2011: 27). Dabei werden neue Elemente oder Kategorien von der einen in die andere

Sprache übernommen, ohne dass in der übernehmenden Sprache bereits existierende Elemente oder Kategorien ersetzt werden (Trudgill 2011: 27).

Kusters (2003) untersucht in seiner Dissertation den möglichen Zusammenhang zwischen dem Wandel der Komplexität in der Verbflexion und dem Wandel einer Sprachgemeinschaft von Typ 1 zu Typ 2. Der Typ 1 hat vor allem L1-Lerner, ist sprecher-orientiert, d.h., der Hörer kann vermuten, was gesagt wird, und Sprecher sowie Hörer verfügen über ein großes gemeinsames Wissen. Des Weiteren hat die Sprache in Gemeinschaften des Typs 1 eine symbolische Funktion: Ausdruck von Identität, ästhetische Funktion etc. (Kusters 2003: 41). Typ-2-Sprachgemeinschaften sind genau das Gegenteil: viele L2-Lerner, kommunikative Funktion (primäres Ziel ist der Austausch von Informationen) und Hörer-orientiert, d.h. die Bedürfnisse des Hörers (z.B. klare, explizite Artikulation) stehen im Vordergrund (Kusters 2003: 41). Die Komplexität der Verbflexion wird definiert als die Phänomene, die für Außenstehende schwierig sind (Kusters 2003: 403). Genauer werden drei Prinzipien herangezogen: Ökonomie, Transparenz und Isomorphie. Werden diese Prinzipien verletzt, gilt die Verbflexion als komplexer: Das Prinzip der Ökonomie ist verletzt, wenn eine Sprache viele Flexionskategorien aufweist; das Prinzip der Transparenz ist verletzt durch Fusion, Fission, Allomorphie und Homonymie; das Prinzip der Isomorphie ist verletzt bei nicht gleichbleibender Affixabfolge in verschiedenen Domänen (Kusters 2003: 403). Untersucht werden Sprachen aus vier typologisch unterschiedlichen Sprachfamilien: Arabisch, skandinavische Sprachen, Quechua und Swahili (Kusters 2003: 403). Für jede dieser Sprachfamilien gibt es Varietäten, deren Sprachgemeinschaft eher dem Typ 1 entspricht und solche, deren Sprachgemeinschaft eher dem Typ 2 entspricht. So kann geprüft werden, ob die Komplexität in der Verbflexion abnimmt, wenn die Sprachgemeinschaft sich von Typ 1 zu Typ 2 wandelt (Kusters 2003: 45). Kurz zusammengefasst kann Kusters (2003) zeigen, dass eine Sprache ökonomischer wird, wenn ihre Sprachgemeinschaft von einem Typ 1 zu einem Typ 2 wird: „[...] [T]he number of categories and category combinations indeed decreases correspondingly in these communities“ (Kusters 2003: 357). Dies hat zur Folge, dass diese Sprachen auch transparenter werden (Kusters 2003: 357). Im Gegensatz dazu scheint Isomorphie keine große Rolle zu spielen (Kusters 2003: 357). Es kann also ein Zusammenhang zwischen dem Typ einer Sprachgemeinschaft und den Phänomenen, die in der Verbflexion vorkommen, also der Komplexität der Verbflexion, beobachtet werden.

McWhorter (2011) vergleicht in seinem ausführlichen und viel beachteten Aufsatz die strukturelle Komplexität einer Kreolsprache (Saramaccan) mit zwei alten Sprachen, wobei die eine (Tses) eher einen synthetischen, die andere (Lahu) eher einen analytischen Sprachbau aufweist. Komplexität wird durch Überspezifikation definiert, d.h. die overte Markierung phonetischer, morphologischer, syntaktischer und semantischer Unterscheidungen, die über die kommunikativen Notwendigkeiten hinausgehen (McWhorter 2001: 125). Bestimmt wird der Grad an Überspezifikation mit der Beschreibungslänge eines Systems: „[...] [S]ome grammars might be seen to require lengthier descriptions in order to characterize even the basics of their grammar than others“ (McWhorter 2001: 134–135). McWhorter (2011) geht dabei aber qualitativ und nicht quantitativ vor, d.h., er macht eine Analyse zu mehr oder weniger Markierung, es werden jedoch für die verschiedenen linguistischen Ebenen keine Zahlen ermittelt und am Schluss miteinander verrechnet. Untersucht wird Überspezifikation

in vier Bereichen. Erstens ist das phonologische System komplexer, je mehr markierte Phoneme das System aufweist (McWhorter 2001: 135). Mit markiert sind selten vorkommende Phoneme gemeint (z.B. Klicks, hintere gerundete Vokale etc.), wobei es hier nicht um komplexe Artikulation, sondern um Implikation geht: Eine Sprache mit markierten Phonemen hat auch unmarkierte und ist somit komplexer als eine Sprache mit unmarkierten Phonemen (McWhorter 2001: 135). Zweitens wird die Syntax komplexer, je mehr Regeln prozessiert werden müssen, wie beispielsweise Asymmetrien zwischen Matrix- und Nebensatz (z.B. Verbzweit-/Verbendstellung), Ergativ/Absolutiv und Nominativ/Akkusativ in derselben Sprache usw. (McWhorter 2001: 136). Drittens wird die Grammatik komplexer, wenn feine semantische und/oder pragmatische Unterscheidungen overt kodiert werden und dieser Ausdruck grammatikalisiert ist (McWhorter 2001: 136). Ein Beispiel hierfür gibt Koasati (Muskogee-Sprache, Nordamerika), das spezielle pronominale Affixe hat, die nur mit stativen Verben verwendet werden (McWhorter 2001: 137). Viertens führt die Flexionsmorphologie generell zu Überspezifikation. Hier werden vor allem drei Bereiche hervorgehoben. A) Morphophonologie und Suppletion: Die Morphophonologie verursacht Prozesse, die phonetisch nur wenig vorausgesagt werden können (z.B. Umlaut) und Suppletion (McWhorter 2001: 137). B) Allomorphie in der Flexion und arbiträre Allomorphie: Zu Ersterem gehören Unterscheidungen zwischen Substantiv- oder Verbklassen, die von der Flexion kodiert werden; zum zweiten Flexion, die mit der Wurzel gelernt und gespeichert werden muss (z.B. Aspektpaare im Russischen) (McWhorter 2001: 138). C) Die Markierung von Kongruenz erhöht ebenfalls die strukturelle Komplexität (McWhorter 2001: 138). In einem ersten Schritt vergleicht McWhorter (2001) die Kreolsprache Saramaccan, die Ende des 17. Jh. entstanden ist, mit der synthetischen Sprache Tsez, eine nakho-daghestanische Sprache aus dem Nordkaukasus. Dabei beobachtet er, dass Tsez häufiger als Saramaccan Überspezifikation zeigt und erklärt dies wie folgt: „[...] [B]y virtue of the fact that they were born as pidgins, and thus stripped of almost all features unnecessary to communication, and since then have not existed as natural languages for a long enough time for diachronic drift to create the weight of “ornament” that encrusts older languages“ (McWhorter 2001: 125). Man könnte also annehmen, dass der Unterschied in der Komplexität zwischen analytischen Sprachen und Kreolsprachen nicht so groß ausfällt (McWhorter 2001: 143–144). Dazu wird Saramaccan mit Lahu verglichen. Das Resultat des Vergleichs zeigt genau das Gegenteil von dem, was man erwarten würde: „It is demonstrated that this complexity differential remains robust even when creoles are compared with older languages lacking inflection [...]“ (McWhorter 2001: 125). Zusammengefasst kann man also festhalten, dass die strukturelle Komplexität einer Sprache mit der Geschichte ihrer Entstehung und ihres Wandels zusammenhängt.

Szmrecsanyi & Kortmann (2009) untersuchen die morphologische/morphosyntaktische Komplexität vieler unterschiedlicher englischer Varietäten, die in folgende Typen eingeteilt werden können: traditionelle L1-Varietäten mit wenig Sprachkontakt (z.B. East Anglia English), L1-Varietäten mit viel Sprachkontakt (z.B. Australian English), Englisch basierte Pidgins und Kreols (z.B. Tok Pisin) und L2-Varietäten (z.B. Hong Kong English) (Szmrecsanyi & Kortmann 2009: 64–65). Die strukturelle Komplexität wird anhand von vier Parametern definiert und gemessen: Ornamentale Regeln, Schwierigkeiten beim L2-Erwerb, Grammatizität/Redundanz, Irregularitäten (Szmrecsanyi & Kortmann 2009: 64–65).

Ornamentale Regeln sind solche, die Unterscheidungen und Asymmetrien hinzufügen, ohne dass sie einen kommunikativen oder funktionalen Vorteil bringen, z.B. Genus (Szmrecsanyi & Kortmann 2009: 68). Die Schwierigkeit beim L2-Erwerb wird durch den Grad gemessen, zu dem eine bestimmte Varietät nicht jene Phänomene aufweist, „[...] that L2 acquisition research has shown to recur in interlanguage varieties“ (Szmrecsanyi & Kortmann 2009: 69). Dazu zählen beispielsweise die fehlende Markierung der Vergangenheit bei regelmäßigen Verben oder die fehlende Inversion. Die Merkmale dieser zwei Parameter sind binär, d.h., entweder kommen sie vor oder nicht (Liste in Szmrecsanyi & Kortmann 2009: 69–69, 70–71). Die Komplexität der vier Typen von Varietäten wird folglich durch das durchschnittliche Vorkommen dieser Merkmale berechnet. Die Datengrundlage hierfür bildet der World Atlas of Morphosyntactic Variation in English (Szmrecsanyi & Kortmann 2009: 65–66). Die Datengrundlage für die Parameter Grammatizität und Irregularität sind verschiedene Korpora, die jedoch keine Pidgins und Kreols beinhalten (Szmrecsanyi & Kortmann 2009: 67). Dies ermöglicht, die strukturelle Komplexität durch die Textfrequenz zu messen. Grammatizität/Redundanz beinhaltet die synthetische Grammatizität, d.h. die gebundenen grammatischen Morpheme und die analytische Grammatizität, d.h. die freien grammatischen Morpheme (Szmrecsanyi & Kortmann 2009: 71–72). Für die synthetische Grammatizität wird ein Syntheseindex berechnet: Die Datenbasis bilden 15.000 Tokens und der Syntheseindex ist die Prozentzahl der gebundenen grammatischen Morpheme pro 1.000 Tokens (Szmrecsanyi & Kortmann 2009: 72). Auf die gleiche Weise wird der Analyseindex berechnet und die Komplexität der Grammatizität bildet die Summe des Synthese- und Analyseindex (Szmrecsanyi & Kortmann 2009: 72). Die Irregularität wird ebenfalls durch die Textfrequenz der gebundenen grammatischen Morpheme gemessen, jedoch getrennt nach regelmäßigen und unregelmäßigen Allomorphen (Szmrecsanyi & Kortmann 2009: 74). Genauer wird ein Transparenzindex ermittelt, und zwar durch den Anteil (in Prozent) der regelmäßigen Allomorphe an allen gebundenen grammatischen Allomorphen (Szmrecsanyi & Kortmann 2009: 74). Es sollen nun kurz die wichtigsten Resultate zusammengefasst werden. Am meisten ornamentale Regeln haben die traditionellen L1-Varietäten, gefolgt von L1-Kontaktvarietäten, Pidgins/Kreols, L2-Varietäten, d.h. „[...] ornamental complexity is clearly a function of the degree of contact [...]“ (Szmrecsanyi & Kortmann 2009: 69). Die Resultate des Parameters L2-Schwierigkeiten zeigen Folgendes (von am meisten zu am wenigsten L2-Merkmale): Pidgins/Kreols > L1-Kontaktvarietäten > traditionelle L1-Varietäten > L2-Varietäten. Diese Resultate sind zu erwarten, jedoch mit Ausnahme der L2-Varietäten, wofür eine Erklärung im Parameter Grammatizität gefunden wird (Szmrecsanyi & Kortmann 2009: 71). Den höchsten Grammatizitätsindex zeigen die traditionellen L1-Varietäten, danach kommen die L1-Kontaktvarietäten und schließlich die L2-Varietäten (Szmrecsanyi & Kortmann 2009: 73). Erstens zeigt dies, dass „[...] a history of contact and adult language learning can eliminate certain types of redundancy [...]“ (Szmrecsanyi & Kortmann 2009: 73). Zweitens ziehen L2-Sprecher nicht die einfacheren den komplexeren Merkmale vor, sondern bevorzugen Null-Markierung. Dies erklärt, weshalb die L2-Varietäten besonders wenige Merkmale aufweisen, die als L2-Merkmale identifiziert wurden (Szmrecsanyi & Kortmann 2009: 73). Drittens ist schließlich besonders interessant, dass es zwischen dem Analyse- und dem Syntheseindex keine Ausgleichstendenzen gibt bzw. dass diese Indizes sogar positiv miteinander korrelieren (Szmrecsanyi & Kortmann 2009: 74). Auch der vierte Parameter weist erwartete Resultate auf. Am transparentesten sind die L2-Varietäten, gefolgt

von den L1-Kontaktvarietäten und zuletzt von den traditionellen L1-Varietäten (Szmrecsanyi & Kortmann 2009: 75). Dies zeigt, dass ausgeprägter Sprachkontakt, das bedeutet vor allem der Spracherwerb von Erwachsenen, dazu führt, Unregelmäßigkeiten abzubauen (Szmrecsanyi & Kortmann 2009: 75). Zusammengefasst erweist sich also, „[...] that variety type is a powerful predictor of complexity variance [...]“, weshalb Sprachkontakt bezüglich des Grads struktureller Komplexität eine zentrale Rolle spielt (Szmrecsanyi & Kortmann 2009: 76).

Maitz & Németh (2014) übernehmen den Forschungsaufbau von Szmrecsanyi & Kortmann (2009) und übertragen ihn auf deutsche Varietäten. Die untersuchten Varietäten sind Zimbrisch (traditionelle L1-Varietät), die deutsche Standardsprache (L1-Kontaktvarietät), Kiche Duits und Unserdeutsch (Pidgin/Kreol) (Maitz & Németh 2014: 7–9). Strukturelle Komplexität wird durch die Parameter Synthetizität, Analytizität, Grammatizität und Irregularität definiert (Maitz & Németh 2014: 6–7). Die Datengrundlage bilden gesprochensprachliche Korpora. Daraus werden 3.000 Wörter zufällig ausgewählt, welche in drei Subsamples à 1.000 Wörter eingeteilt sind (Maitz & Németh 2014: 9). Synthetizität wird durch den Anteil der Wörter am Korpus gemessen (in Prozent), die durch gebundene grammatische Morpheme markiert sind (Maitz & Németh 2014: 10). Dazu gehören die Kasus- und Numerusmarkierung am Nomen, Pronomen und Artikel, die Numerus- und Genusmarkierung und die Komparativformen des Adjektivs, die Tempusmarkierung und die Markierung der Person am Verb (Maitz & Németh 2014: 10). Analytizität wird durch den Anteil an Funktionswörtern am Korpus ermittelt, z.B. Determinierer, Konjunktionen, Auxiliärverben etc. (Maitz & Németh 2014: 11–12). Die Grammatizität besteht aus der Summe der Synthetizität und der Analytizität (Maitz & Németh 2014: 12). Schließlich wird die Irregularität anhand der durchschnittlichen Textfrequenz von unregelmäßigen Markern berechnet, wozu z.B. Stammmodifikationen und Suppletion gehören (Maitz & Németh 2014: 12–13). Die Parameter Synthetizität, Analytizität und Grammatizität zeigen dieselben Resultate: Die höchste Komplexität zeigt Zimbrisch (traditionelle L1-Varietät), gefolgt von der Standardsprache (L1-Kontaktvarietät) und zuletzt von Kiche Duits/Unserdeutsch (Pidgin/Kreol), wobei der Unterschied nur zwischen Zimbrisch und der Standardsprache sowie zwischen Zimbrisch und Kiche Duits/Unserdeutsch signifikant ist, jedoch nicht zwischen der Standardsprache und Kiche Duits/Unserdeutsch (Maitz & Németh 2014: 11–13). Gleiches gilt für den Parameter Irregularität, nur dass hier auch der Unterschied zwischen der Standardsprache und Kiche Duits/Unserdeutsch signifikant ist (Maitz & Németh 2014: 14). Des Weiteren wird eine positive Korrelation zwischen Synthetizität und Analytizität gefunden, was gegen die oft angenommene Ausgleichstendenz zwischen synthetischen und analytischen Phänomenen spricht (Maitz & Németh 2014: 15). Auch eine positive Korrelation zwischen Grammatizität und Irregularität kann beobachtet werden (Maitz & Németh 2014: 15). Folglich ist ein hoher Grad an Irregularität typisch für morphosyntaktisch komplexe Varietäten (Maitz & Németh 2014: 17). Es kann also festgehalten werden, dass Sprachkontakt, d.h. Spracherwerb von Erwachsenen, zu Simplifizierungen in der Grammatik führt. Maitz & Németh (2014) geben aber ebenfalls zu bedenken, dass auch weitere Faktoren die Morphologie vereinfachen können, wie beispielsweise phonologischer Wandel (z.B. Schwächung der Nebensilbe) (Maitz & Németh 2014: 20). Gleichzeitig können jedoch

ebenfalls Faktoren wirken, die Simplifizierung hemmen, wie z.B. die Attitüden gegenüber präskriptiven Normen (Maitz & Németh 2014: 21).

Auch **Schreier (2016)** geht davon aus, dass Sprachkontakt die strukturelle Komplexität einer Sprache oder einer Varietät beeinflussen kann. Im Gegensatz jedoch zu den bis hier erörterten Studien kann er zeigen, dass Kontakt nicht nur zu Simplifizierung sondern auch zu Komplexifizierung führt, und zwar in derselben Varietät zur selben Zeit (Schreier 2016: 139, 148–151). Auch die (sozio-)linguistischen Faktoren differenziert er genauer, wobei es sich um folgende handelt: a) die Eigenschaften der Systeme, die miteinander in Kontakt stehen; b) das soziale Fundament des Kontakts, wozu auch demografische Daten und die Besiedlung gehören; c) die Intensität des Kontakts, wie z.B. die Stärke von Substrateffekten (Schreier 2016: 139). Es geht vor allem darum, „[...] to do justice to the multifaceted and interwoven (social and linguistic) phenomena of contact-induced change“ (Schreier 2016: 139). Vor allem kann Schreier (2016) zeigen, dass die Dichotomie – Varietäten mit viel Kontakt werden einfacher, während Varietäten mit wenig Kontakt komplexer werden – zu kurz greift. Als erstes wird anhand der Reduktion von Konsonantenclustern dargestellt, dass nicht die Intensität des Sprachkontakts entscheidend ist, sondern die systemischen und typologischen Unterschiede zwischen den Sprachen im Kontakt: Der starke Kontakt von zwei Varietäten, die beide komplexe Konsonantencluster aufweisen, hat kaum einen Effekt, während ein weniger starker Kontakt mit einer Varietät mit einfacheren Konsonantenclustern einen höheren Effekt hat (Schreier 2016: 144). Dass starker Kontakt zu Simplifizierungen führt, trifft folglich nur auf unterschiedliche Systeme zu, nicht aber, wenn die Systeme große Ähnlichkeiten aufweisen (Schreier 2016: 144–145). Zweitens zeigt Schreier (2016), dass das System der Personalpronomen von Varietäten mit wenig Kontakt relativ regelmäßig, also wenig komplex ist und dass jenes von Varietäten mit viel Kontakt höhere Komplexität aufweist (Schreier 2016: 145–147). Beispielsweise hat Tok Pisin in der 1. Person Plural zusätzlich die Unterscheidung inklusiv/exklusiv grammatikalisiert (Schreier 2016: 147). Drittens ist festzustellen, dass Tristan da Cunha English weder eine rein *low-contact* noch eine rein *high-contact* Varietät ist und gleichzeitig Simplifizierung und Komplexifizierung aufweist. Je nachdem welches Zeitfenster man sich anschaut, handelt es sich bei Tristan da Cunha English um eine Varietät mit viel oder mit wenig Kontakt (Schreier 2016: 148). Beispiele für Simplifizierung sind die Ausgleiche im Paradigma von *sein* (*is* vs. *was* unabhängig von Person und Numerus) oder die hohe Rate an Vereinfachung komplexer Konsonantencluster (Schreier 2016: 149–150). Komplexifizierung zeigt sich z.B. darin, dass der dentale Frikativ vier Realisierungen hat: dentale Frikative, labiodentale Frikative und labiodentale Plosive sowie Sibilanten (Schreier 2016: 150). Schließlich kann noch angefügt werden, dass Tristan da Cunha English Archaismen erhalten hat und gleichzeitig aber auch Innovationen aufweist (Schreier 2016: 150–151). Aufgrund dieser Resultate kritisiert Schreier (2016) besonders drei Aspekte der bisherigen Modelle. Erstens ist es schwierig, Varietäten in die binäre Unterscheidung zwischen Varietäten mit viel Kontakt und Varietäten mit wenig Kontakt einzuteilen (Schreier 2016: 151–152). Zweitens berücksichtigen die Modelle die Diachronie zu wenig, denn „[...] isolation effects may override contact effects over time“ (Schreier 2016: 152). Drittens können Komplexifizierung und Simplifizierung gleichzeitig auftreten (Schreier 2016: 153). Schreier (2016) schlägt ein Modell vor, das wenig Kontakt und viel Kontakt als die beiden Pole auf einem Kontinuum ansetzt (Schreier 2016: 153–154).

Dieses Modell trägt der Tatsache Rechnung, dass eine Varietät komplexe und einfache Phänomene gleichzeitig aufweisen kann.

2.2.4. Geografische Isolation

Im vorangehenden Kapitel wurden einige Studien skizziert, in denen mit dem Begriff Isolation, wenn er verwendet wird, vorwiegend soziale Isolation gemeint ist, d.h. Sprachgemeinschaften mit wenig Sprachkontakt, wenig L2-Sprechern etc. Vor allem bei Jakobson (1929), Braunmüller (1984) und Trudgill (u.a. 2011) impliziert Isolation jedoch meistens auch geografische Isolation. In diesem Kapitel sollen drei Untersuchungen vorgestellt werden, die sich explizit mit geografischer Isolation beschäftigen, nämlich Nichols (1992 und 2016) und Garzonio (2016). Obwohl diese Studien erst Ende des 20. Jh./Anfang des 21. Jh. entstanden sind, weist bereits Jakobson (1929) darauf hin, dass die ukrainischen Dialekte mit einem großen Vokalinventar sich eher an der Peripherie des ukrainischen Sprachgebiets befinden: „Cette différence [entre les systèmes de voyelles] est due, en premier lieu à la tendance conservatrice qui est caractéristique des parlers de la périphérie, et en second lieu à des différences fonctionnelles“ (Jakobson 1929: 73, ausführliches Zitat in Kap. 2.2.1.).

Die wichtigsten Aspekte aus **Nichols (1992)** für das Ziel dieser Arbeit wurden in Baechler und Seiler (2016) zusammengefasst. Die anschließenden Ausführungen folgen dieser Zusammenfassung. Nichols ist wahrscheinlich die erste Linguistin, die die mögliche Korrelation zwischen geografischer Isolation und struktureller Komplexität explizit untersucht hat. Ihr wichtigstes Ziel ist, typologische Eigenschaften und Geografie zu korrelieren: „In describing the distribution of types and typological features we can often make active use of geography as a predictive factor. This means viewing the languages of a region as a population and demonstrating a correlation between the location or type of the region and the distribution of traits within the population or between populations“ (Nichols 1992: 12). Nichols (1992) unterscheidet u.a. zwischen *residual* und *spread Zones*, wobei ersterer als isoliertes Areal und letzterer als nicht isoliertes Areal bezeichnet werden kann. *Spread Zones* sind wie folgt charakterisiert:

- (1) Little genetic diversity [...].
- (2) Low structural diversity.
- (3) The language families present in the spread zone are shallow.
- (4) Rapid spread of languages or language families and consequent language succession.
- (5) Classic dialectal-geographical area with innovating center and conservative periphery [...].
- (6) No net long-term increase in diversity. A spread zone is a long-lasting phenomenon, but it preserves little linguistic evidence of its history.
- (7) The spreading language serves as a lingua franca for the entire area or a large part of it.

(Nichols 1992: 16–17)

Im Gegensatz zu den *spread Zones* weisen die *residual Zones* folgende typischen Charakteristika auf:

- (1) High genetic density [...].
- (2) High structural diversity [...].
- (3) The language families [...] are deep.
- (4) No appreciable spread of languages or families. No language succession.
- (5) No clear center of innovation [...]
- (6) Accretion of languages and long-term net increase in diversity. [...].
- (7) No lingua franca [...] for the entire area; local bilingualism or multilingualism is the main means of inter-ethnic communication.

(Nichols 1992: 21)

Das wichtigste Resultat von Nichols' Studie bezüglich der möglichen Korrelation zwischen Isolation und struktureller Komplexität ist, dass „[r]esidual zones show relatively high complexity, equal to or greater than that of their respective continents. Spread zones show somewhat lower average complexity, equal to or lower than that of their respective continents“ (Nichols 1992: 192). Nichols (1992) nimmt aber auch an, dass es innovative und konservative Gegenden innerhalb der *residual* und *spread Zones* gibt. In einer *spread Zone* kann ein innovatives Zentrum und eine konservative Peripherie gefunden werden (Nichols 1992: 17). In bergigen Regionen, die als *residual Zones* gelten können, gibt es Innovationen eher an der Peripherie, d.h. im Tal oder in der Ebene, und Archaismen im Zentrum, d.h. im Gebirge (Nichols 1992: 14).

Dass *spread* und *residual Zones* tatsächlich innovative und konservative Areale aufweisen, kann Nichols (2016) in einer weiteren Untersuchung zeigen. Dazu werden Sprachen aus dem Ostkaukasus und aus der eurasischen Steppe analysiert. Bei den Sprachen aus dem Ostkaukasus handelt es sich um solche aus der nakho-daghestanischen Sprachfamilie (Nichols 2016: 120–121). Aus der eurasischen Steppe werden 22 Sprachen aus folgenden Sprachfamilien herangezogen, wobei einige Sprachen keiner Sprachfamilie zugeordnet werden können (es werden nur die Sprachfamilien bzw. die isolierten Sprachen aufgezählt): Turksprachen, Mongolisch, Tungusisch, Tibetisch, Koreanisch, Japanisch, Ainu, Nivkh, Tschuktschisch-Kamtschadalisch, Eskimo-Aleutisch, Jukagir, Jenissej, östliches Uralisch (Nichols 2016: 129). Eingeteilt werden die Sprachen in zwei Kategorien, und zwar in sich ausbreitende Sprachen (mit vielen L2-Lernern, Lingua Franca) und in isolierte, sich nicht ausbreitende Sprachen. Gemessen wird die strukturelle Komplexität durch die Größe des

Inventars (d.h. die Anzahl der Elemente, z.B. die Anzahl der Phoneme, Flexionskategorien, Wortabfolgen etc.) und durch den Grad der Opazität (z.B. Suppletion, Sandhi, Mehrfachausdruck, Synthese etc.) (Nichols 2016: 118). Die Größe des Inventars wird auf verschiedenen linguistischen Ebenen ermittelt, nämlich Phonologie, Synthese, Klassifizierer, Syntax und Lexikon, und Opazität durch Genus und Deklination (Nichols 2016: 137). Eine genaue Liste der untersuchten linguistischen Variablen ist aus Nichols (2016: 137) zu entnehmen. Quantifiziert wird auf zwei Arten, und zwar je nachdem, ob es sich um eine binäre Variable handelt oder nicht. Auf der phonologischen Ebene kann beispielsweise die Anzahl der kontrastiven obstruenten Arten der Artikulation einfach gefunden werden, während Töne eine binäre Variable darstellen, die mit +1 für das Vorhandensein dieser Variable zu Buche schlägt und mit +0 für das nicht Vorhandensein (Nichols 2016: 137). Für den Ostkaukasus kann Nichols (2016) zeigen, dass Opazität mit isolierten Sprachen korreliert, während Transparenz und eine geringe Inventargröße in sich ausbreitenden Sprachen zu beobachten sind (Nichols 2016: 129). Da es sich um ein gebirgiges Gebiet handelt, kann Isolation durch geografische Höhe quantifiziert werden (Nichols 2016: 129). Diese Resultate sind aufgrund des soziolinguistischen Kontextes dieser Region zu erwarten, die kurz erörtert werden. Die Sprachen, die in der Ebene oder am Fuße eines Berges situiert sind, breiten sich bergaufwärts aus und weisen deshalb viele L2-Sprecher auf (Nichols 2016: 123–124). In den Bergdörfern hingegen sind enge Netzwerke und kaum L2-Lerner der isolierten Sprache zu beobachten (Nichols 2016: 123–124). Neben den nakho-daghestanischen Sprachen des Ostkaukasus untersucht Nichols (2016) auch unterschiedliche Sprachen der eurasischen Steppe. In diesen Sprachen der eurasischen Steppe weisen sich ausbreitende Sprachen eine geringe Inventargröße auf, während isolierte Sprachen an der Peripherie des Gebiets einen hohen Grad an Opazität haben (Nichols 2016: 129–130). Die eurasische Steppe kann eine lange Geschichte an Ausbreitung unterschiedlicher Sprachen vorweisen, verschiedene Sprachen haben zu unterschiedlichen Zeiten sich ausgebreitet und überlagert. Sie hatten also immer viele L2-Lerner, was zur Simplifizierung der Grammatik führt (Nichols 2016: 118–119). Den Sprachen an der Peripherie (besonders im Norden) fehlen jüngste Ausbreitungsphasen, oder sie haben nie welche gehabt, d.h., diese Sprachen verfügten über längere Zeitabschnitte, in denen sich natürliche Komplexifizierung durch L1-Übertragung entwickeln konnte (Nichols 2016: 118–119). Es kann also festgehalten werden, dass geringe Inventargröße vor allem in sich ausbreitenden Sprachen gefunden werden kann, die zur Kommunikation zwischen verschiedenen Gesellschaften dienen, und hohe Opazität in isolierten Sprachen, die generell nicht als L2 gelernt werden (Nichols 2016: 132–133).

Garzonio (2016) untersucht die syntaktische Komplexität von Entscheidungsfragesätzen sowie w-Fragesätzen, und zwar in den folgenden alpinen italo-romanischen Dialekten: Monnese, Bellunese und Mendrisiotto (Garzonio 2016: 95). Zum Sample gehört auch das übliche System nicht alpiner norditalienischer Dialekte (Garzonio 2016: 95). Alle drei Dialekte liegen im alpinen Gebiet und können im Gegensatz zu vielen anderen norditalienischen Dialekten als isoliert gelten, wobei Monnese noch etwas isolierter ist als Bellunese und Mendrisiotto (Garzonio 2016: 95, 102). Die syntaktische Komplexität wird anhand zweier Parameter gemessen: Komplexität der Derivation und die Anzahl freier Varianten für dieselbe Funktion (= Optionalität) (Garzonio 2016: 98–99). Die Komplexität der Derivation wird anhand der Anzahl Move- und Mergeoperationen quantifiziert: je mehr

Move- und Mergeoperationen, desto komplexer die Derivation (Garzonio 2016: 98–99). Die Resultate sind in der folgenden Tabelle zusammengefasst:

Tabelle 2.1: Resultate syntaktischer Komplexität, Move/Merge (Garzonio 2016: 111–112)

Komplexität der Derivation:		
	Entscheidungsfragesätze	w-Fragesätze
norditalienische Dialekte	1/1	2/1
Monnese	1/1	2/1
Bellunese	1/1	1/1
Mendrisiotto	0/1	1/1
Optionalität:		
	w-Fragesätze	
norditalienische Dialekte	1	
Monnese	3	
Bellunese	1	
Mendrisiotto	5 (3)	

Bezüglich der Optionalität zeigt sich, dass Monnese und Mendrisiotto deutlich komplexer sind. Dies hat vor allem damit zu tun, dass in diesen Dialekten das w-Element verdoppelt werden und dann links und/oder rechtsperipher stehen kann (Garzonio 2016: 105, 108–111). Mendrisiotto weist zudem im Gegensatz zu den übrigen Varietäten des Samples in den w-Elementen drei verschiedene Formen auf: *cusè*, *cusa*, *sa* (Garzonio 2016: 109). Die Komplexität der Derivation betreffend sind die nicht alpinen norditalienischen Dialekte und Monnese komplexer als Bellunese und Mendrisiotto. Garzonio kann jedoch zeigen, dass in Bellunese und Mendrisiotto die w-Fragesätze mit einer speziellen Semantik höhere Komplexität aufweisen, nämlich zwei Move- und zwei Mergeoperationen (Garzonio 2016: 108, 111). Schließlich sind zwei interessante Beobachtungen hervorzuheben. Erstens hat Monnese eine *tun*-Periphrase grammatikalisiert, die jedoch in der Derivation nicht zu höherer Komplexität führt (Garzonio 2016: 103–104). Isolation hat in diesem Fall also nicht höhere syntaktische Komplexität hervorgebracht, sondern eine andere Syntax (Garzonio 2016: 113). Zweitens weist Medrisiotto in der Derivation die geringste Komplexität, aber in der Optionalität die höchste Komplexität auf (Garzonio 2016: 111–114). Zusammengefasst kann also Folgendes festgehalten werden: Erstens sind isolierte Dialekte nicht komplexer als nicht isolierte bezüglich der derivationellen Komplexität, jedoch können sie einen hohen Grad an Optionalität aufweisen; zweitens variiert die derivationelle Komplexität zwischen verwandten Varietäten nur wenig; drittens können sich Fragesatztypen in ihrer Komplexität unterscheiden (am extremsten im Mendrisiotto) (Garzonio 2016: 114).

2.2.5. Arten von Komplexität

Hier soll nun zuerst die etablierte Unterscheidung zwischen absoluter und relativer Komplexität vorgestellt werden, die von Miestamo (2008) vorgeschlagen wird. Danach wird eine detailliertere Einteilung der Typen struktureller Komplexität erörtert, die auf den Philosophen Rescher (1998) zurückgeht und von Miestamo, Sinnemäki & Karlsson (2008) für linguistische Phänomene adaptiert wurde. Schließlich werden die Definitionen struktureller Komplexität der vorangehenden Kapitel zusammengefasst und kategorisiert.

Neben lokaler und globaler Komplexität (vgl. Kap. 2.1.2.) unterscheidet Miestamo (2008) zwischen absoluter und relativer Komplexität. Im relativen Ansatz wird Komplexität durch den Aufwand oder Schwierigkeit für den Sprecher definiert, d.h., „how difficult a phenomenon is to process (encode/decode) or learn. The more costly or difficult a linguistic phenomenon is, the more complex it is [...]“ (Miestamo 2008: 25). Es stellt sich also immer die Frage, für wen ein linguistisches Phänomen schwierig oder aufwändig ist, wie z.B. Hörer, Sprecher, L1- oder L2-Lerner. Dabei ist zu beachten, dass ein linguistisches Phänomen für eine Gruppe (z.B. Hörer) Aufwand oder Schwierigkeiten verursacht, während dasselbe Phänomen für eine andere Gruppe (z.B. Sprecher) einfach ist (Miestamo 2008: 25). Beschäftigt man sich mit absoluter Komplexität, steht ausschließlich das linguistische System im Vordergrund. Die Grundidee dahinter ist, je mehr Elemente ein System hat, desto komplexer ist es (Miestamo 2008: 24). Um die Idee der Anzahl Elemente genereller zu fassen, greift Miestamo (2008) auf Informationstheorien zurück (z.B. Shannon-, Kolmogorov-Komplexität) und schlägt, basierend auf Dahl (2004), folgende Definition absoluter Komplexität vor:

[...] [T]he complexity of a linguistic phenomenon may be measured in terms of the length of the description of that phenomenon; the longer a description a phenomenon requires, the more complex it is [...] A less complex phenomenon can be compressed to a shorter description without losing information. On a high level of abstraction we may say that we are still dealing with the number of parts in a system, but these parts are now the elements that constitute the description of the system. (Miestamo 2008: 24–25)

In Arbeiten zur strukturellen Komplexität sind bereits unzählige Übersichten zusammengestellt worden, die resümieren, was genau mit struktureller Komplexität gemeint ist und welche Arten struktureller Komplexität in Sprachen vorgefunden werden (z.B. Szendrői & Kortmann 2012: 10–12). Im Gegensatz dazu gibt Rescher (1998) in seiner philosophischen Auseinandersetzung mit dem Thema Komplexität eine breitere Definition und Klassifikation der Arten von Komplexität. Komplexität definiert er folgendermaßen:

Complexity is first and foremost a matter of the number and variety of an item's constituent elements and of the elaborateness of their interrelational structure, be it organizational or operational. Any sort of system or process – anything that is a structured whole consisting of interrelated parts – will be to some extent complex. Accordingly, all manner of things can be more or less complex: natural objects (plants or river systems), physical artifacts (watches or sailboats), mind-engendered processes (languages or instructions), bodies of knowledge, and so on. (Rescher 1998: 1)

Des Weiteren unterscheidet Rescher (1998) zwischen unterschiedlichen Typen von Komplexität (Rescher 1998: 8–16). Diese Unterscheidung wurde von Miestamo, Sinnemäki und Karlsson (2008) übernommen und für die Kategorisierung linguistischer Phänomene angepasst (Miestamo, Sinnemäki & Karlsson 2008: VIII–IX). Davon erstellt Sinnemäki (2011) eine modifizierte Version, die in Tabelle 2.2 wiedergegeben ist und welche weiter unten anhand der Tabelle 2.3 genauer erörtert wird:

Tabelle 2.2: Arten der Komplexität (Rescher 1998: 9; Sinnemäki 2011: 23)

<p>1. Epistemic modes</p> <p><u>A. Formulaic complexity</u></p> <p>a. Descriptive complexity: length of the account that must be given to provide an adequate description of a given system.</p> <p>b. Generative complexity: length of the set of instructions that must be given to provide a recipe for producing a given system.</p> <p>c. Computational complexity: amount of time and effort involved in resolving a problem.</p> <p>2. Ontological modes</p> <p><u>A. Compositional complexity</u></p> <p>a. Constitutional complexity: number of constituent elements (e.g., in terms of the number of phonemes, morphemes, word, or clauses).</p> <p>b. Taxonomic complexity (or heterogeneity): variety of constituent elements, that is, the number of different kinds of components (e.g., tense-aspect distinctions, clause types).</p> <p><u>B. Structural complexity</u></p> <p>a. Organizational complexity: variety of ways of arranging components in different modes of interrelationship (e.g., phonotactic restrictions, variety of distinctive word orders).</p> <p>b. Hierarchical complexity: elaborateness of subordination relationships in the modes of inclusion and subsumption (e.g., recursion, intermediate levels in lexical-semantic hierarchies).</p> <p>3. Functional complexity</p> <p><u>A. Operational complexity</u>: variety of modes of operation or types of functioning (e.g., cost-related differences concerning the production and comprehension of utterances).</p> <p><u>B. Nomic complexity</u>: elaborateness and intricacy of the laws governing a phenomenon (e.g., anatomical and neurological constraints on speech production; memory restrictions).</p>

Vergleicht man diese Arten von Komplexität und Miestamos (2008) Einteilung in absolute und relative Komplexität, kann die epistemische und die ontologische Komplexität der absoluten Komplexität zugeteilt werden und die funktionale Komplexität der relativen Komplexität.

Abschließend soll nun versucht werden, die in diesem Kapitel diskutierten Arbeiten in die Kategorisierung der Komplexitätsarten von Rescher (1998) und Sinnemäki (2011)

einzuordnen. Eine tabellarische Übersicht gibt Tabelle 2.3, die die Komplexitätsarten auslöst, welche in den erörterten Arbeiten keine Rolle spielen, da es das Ziel ist, einen Überblick zu bieten. Deswegen werden linguistische Konzepte und Prinzipien einzelnen Komplexitätsarten zugeordnet, was natürlich eine Verallgemeinerung zur Konsequenz hat.

Alle Studien mit Ausnahme von Sampson (2001) und Garzonio (2016) beruhen mehr oder weniger explizit auf der Grundidee der deskriptiven Komplexität: Je länger die Beschreibung der Grammatik ist, desto komplexer ist die Grammatik. Ausschließlich Garzonio (2016) operationalisiert die generative Komplexität, um die Komplexität einer Sprache zu messen. Dazu zählt er die Anzahl von Move- und Mergeoperationen, wobei es sich um Instruktionen zur Ableitung (also Produktion) wohlgeformter Sätze handelt. Die ontologischen Arten bilden den Schwerpunkt in der bisherigen Beschäftigung mit linguistischer Komplexität. Diese werden unterteilt in kompositionelle und strukturelle Komplexität. Zur kompositionellen Komplexität gehören die Größe des Inventars (konstitutionelle Komplexität) und die Anzahl Unterscheidungen (und deren Kodierung), die in einem System gemacht werden (taxonomische Komplexität). Die Inventargröße wird vorwiegend in der Phonologie untersucht, wie z.B. das Phon-/Phoneminventar (Jakobson (1929), Hay & Bauer (2007), Nichols (2016), Schreier (2016)). Dazu gehört aber auch die Komplexität der Konsonantencluster bei Schreier (2016) sowie die markierten Phoneme bei McWhorter (2001), da es hier um eine Implikation geht: Ein Sprachsystem, das markierte Phoneme hat, hat auch unmarkierte und folglich ein größeres Inventar. Zur Inventargröße kommt bei Nichols (2016) neben der Größe des Phoneminventars die Anzahl unterschiedlicher Präfixtypen dazu. Bei allen anderen Variablen, die Nichols (2016) unter Inventargröße subsummiert, handelt es sich nach Reschers (1998) und Sinnemäki (2011) Einteilung eher um taxonomische Komplexität. Damit ist gemeint, dass je mehr Unterscheidungen (und Kodierungen dieser Unterscheidungen) gemacht werden, desto komplexer ist ein System. Beispielsweise ist ein System, das Genus unterscheidet, komplexer als eines, das Genus nicht unterscheidet, und je mehr Genera ein System unterscheidet, desto komplexer ist es ebenfalls. Dazu gehören die meisten Variablen bei Nichols (2016), die ornamentalen Regeln bei Szmrecsanyi & Kortmann (2009) sowie der Verstoß gegen das Ökonomieprinzip bei Kusters (2003). Auch McWhorters Konzept der Überspezifikation (2001) kann zur taxonomischen Komplexität gerechnet werden und insbesondere sein dritter Parameter, nämlich die overt kodierten feinen semantischen und/oder pragmatischen Unterscheidungen. Schließlich zählen bei Schreier (2016) die Grammatikalisierung von Exklusiv/Inklusiv in den Personalpronomen wie auch die Ausgleiche im Verbalparadigma (Verlust von Person- und Numerusunterscheidung, also Simplifizierung) dazu. In der strukturellen Komplexität wird zwischen der organisationellen und der hierarchischen Komplexität unterschieden. Die organisationelle Komplexität beinhaltet jene Phänomene, die das 1-zu-1-Verhältnis zwischen Form und Bedeutung verletzen (Sinnemäki 2011: 25). Bei Sinnemäki (2009) ist das die Verletzung von Ökonomie und Distinktheit, d.h., eine Funktion (hier Markierung von Agens und Patiens) wird zu viel oder zu wenig markiert. Auch die Grammatizität bei Maitz & Németh (2014) sowie Szmrecsanyi & Kortmann (2009) könnte dazugezählt werden. Die Grammatizität wird durch die Textfrequenz freier und gebundener grammatischer Morpheme gemessen. Damit können jedoch weder Aussagen gemacht werden, ob eine Funktion zu viel oder zu wenig kodiert wird, noch, welche und wie viele Funktionen kodiert werden. Das

Konzept der Grammatizität ist also nur schwer in Reschers Kategorisierung einzuordnen. Des Weiteren gehören zur organisationelle Komplexität alle Phänomene, die unter dem Konzept Opazität/Transparenz zusammengefasst werden können: morphophonologische Regeln, Suppletion, Allomorphie, Homonymie, Fusion, Irregularität etc. Arbeiten dazu stammen von Braunmüller (1984), McWhorter (2001), Kusters (2003), Szmrecsanyi & Kortmann (2009), Maitz & Németh (2014) sowie Nichols (2016). Zur organisationellen Komplexität wird auch die Anzahl unterschiedlicher Wortabfolgen gerechnet (bei McWhorter (2001), Nichols (2016)). Schließlich untersucht Sampson (2001) hierarchische Komplexität, indem die Einbettungstiefe gemessen wird. Wie bereits oben erwähnt, zählen die epistemische und die ontologische Komplexität zur absoluten Komplexität und die funktionale Komplexität zur relativen Komplexität. Zur funktionalen Komplexität gehört die operationelle Komplexität, womit u.a. der Aufwand und die Schwierigkeiten in der Produktion und Rezeption der Sprache durch unterschiedliche Gruppen (z.B. Sprecher, Hörer, L1- oder L2-Lerner) gemeint sind. Dies spielt besonders bei Kusters (2003) eine zentrale Rolle, bei Szmrecsanyi & Kortmann (2009) betrifft dies den zweiten von vier Parametern. Die vorliegende Arbeit beschäftigt sich mit der generativen Komplexität, die genauer durch konstitutionelle, taxonomische und organisationelle Komplexität definiert ist. Genauer wird dies im anschließenden Kapitel erläutert.

Tabelle 2.3: Einordnung der diskutierten Arbeiten in die Komplexitätsarten nach Rescher (1998) und Sinnemäki (2011)

1. Epistemic modes

A. Formulaic complexity

- a. Descriptive complexity: alle (außer Sampson (2001) und Garzonio (2016))
- b. Generative complexity: Garzonio (2016)

2. Ontological modes

A. Compositional complexity

- a. Constitutional complexity: Jakobson (1929), McWhorter (2001), Hay & Bauer (2007), Nichols (2016), Schreier (2016)
- b. Taxonomic complexity (or heterogeneity): McWhorter (2001), Kusters (2003), Sinnemäki (2009), Szmrecsanyi & Kortmann (2009), Maitz & Németh (2014), Nichols (2016), Schreier (2016)

B. Structural complexity

- a. Organizational complexity: Transparenz/Opazität: Braunmüller (1984), McWhorter (2001), Kusters (2003), Sinnemäki (2009), Nichols (2016); Wortabfolge: McWhorter (2001), Nichols (2016)
- b. Hierarchical complexity: Sampson (2001)

3. Functional complexity

- A. Operational complexity: Kusters (2003), Szmrecsanyi & Kortmann (2009)