Innovation and grammaticalization in the emergence of Tok Pisin

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophie
der Ludwig-Maximilians-Universität
München

vorgelegt von

Martin Eberl

aus

Freising

2021

Zusammenfassung

Ziel der Arbeit

Die vorliegende Arbeit analysiert den Ursprung verschiedener grammatikalischer Strukturen in Tok Pisin, der Varietät des Melanesischen Pidgins, das in Papua Neu Guinea als Verkehrssprache fungiert. Durch die Analyse dieser Strukturen wird aufgezeigt, wie Sprecher des Tok Pisin während der kritischen Phase des Entstehens der Sprache die Prozesse der Reanalyse und Grammatikalisierung genutzt haben, um wichtige Strukturen der Sprache zu formen, wie sie noch heute gesprochen wird. Anstatt vor allem auf lexikalisches Material und andere Strukturen der diversen Ursprungssprachen zurückzugreifen, bedienten sich frühe Sprecher der sprachinternen Reanalyse und Innovation, um ein zunehmend komplexes System zu festigen. Indem analysiert wird, wie dieser Prozess vonstattenging, ergeben sich wertvolle Einblicke dahingehend, wie Sprachen, oder, genauer gesagt, ihre Sprecher, mit funktioneller Nachfrage umgehen: zum einen durch die Schaffung neuer Elemente und das Einbringen solcher von außen, zum anderen aber besonders auch durch die Rekombination und Reanalyse bestehender, interner Elemente.

Die Arbeit geht dabei von einer scheinbar einfachen Ursprungsfrage aus: warum ist menschliche Sprache derart regelkonform? Das mag zunächst eine seltsame Frage scheinen, ist aber nähere Betrachtung wert. Als Sprecher einer Sprache sind wir frei dazu in der Lage, jedes mögliche Wort zu erfinden. Wir sind ebenso in der Lage, die Struktur jedes Satzes zu verändern. Es besteht kein physikalisches Limit, welches uns daran hindern würde, die Töne einer Sprache zu verändern (einmal abgesehen von denjenigen, die uns auf die möglichen Töne des menschlichen Sprechapparats beschränken), und es gibt keinen bestimmten Grund, warum wir nicht neue Zeitformen erfinden sollten, mit unseren ganz eigenen verbalen Suffixen oder freistehenden Signalwörtern, die diese formal ausdrücken. Warum haben wir also nicht genausoviele grammatikalische Systeme, wie wir Sprecher einer Sprache haben?
Die Antwort darauf - oder zumindest ein Teil der Antwort - liegt natürlich in

der Tatsache, dass es eben doch Beschränkungen für jedes bestehende Sprachsystem gibt. Eine dieser Beschränkungen, und sicherlich eine der wichtigsten, ist die der Verständlichkeit. Würden wir die Elemente unserer Sprache zu sehr verändern, würden wir von anderen Sprechern nicht länger verstanden werden. Würden wir gezielt Schlüsselelemente einer Sprache verändern, könnten andere Spreche unsere Aussagen entweder gar nicht, oder missverstehen. Die neuartigen, kreativen Elemente, die wir eingeführt haben, würden dann als Fehler abgetan.

Es gibt allerdings durchaus Situationen, in denen Sprecher einer Sprache dazu gezwungen sind, kreative Lösungen zu entwickeln. Entweder fehlt ihnen, oder dem Sprachsystem als Ganzem, eine Möglichkeit, ein gewisses Konzept, sei es lexikalisch oder grammatikalisch, auszudrücken. Daher haben wir Kategorien wie das Wort des Jahres, in dem wir neue, kreative Ausdrücke ehren - oder vor ihnen mit (gespielter) Abscheu zurückschrecken, je nach unser präskriptivistischen Anlage. Und dann gibt es natürlich Änderungen an den grammatikalischen und syntaktischen Systemen einer Sprache über die Dauer ihres Bestehens. Sprache als Ganzes und Sprachen im Einzelnen waren nie starre Konstrukte, sondern haben sich schon immer mit ihren Sprechern weiterentwickelt.

Wir können also kreativ sein, was unsere Sprache angeht - bis zu einem gewissen Punkt. Aber welcher Punkt ist das? Und gibt es Umstände, Punkte in der Entwicklung einer Sprache, die uns zwingen, kreativer zu sein als sonst? Eine mögliche Antwort auf letztere Frage könnten Situationen sein, in denen das Entstehen und die weitere Entwicklung einer Sprache schneller und unter größerem Druck geschehen muss, als dies normalerweise der Fall ist. Solche Argumente wurden in der Vergangenheit für das Entstehen von Kontaktsprachen wie Pidgin- und Kreolsprachen angebracht, denen nachgesagt wird, sie würden sich aufgrund ihrer Rolle als Lingua Franca zwischen Sprechern verschiedener Sprachen schneller entwickeln. Eine solche Sprache ist Tok Pisin, wie es in Papua Neu Guinea gesprochen wird.

Könnten solche Sprachen also Hinweise darauf liefern, ob unsere sprachliche Kreativität unter solchen Umständen ihre Fesseln abwirft? Und was würde dies über die Strukturen und Muster aussagen, die solche Sprachen entwickeln? Würden sie sich signifikant von typologischen Mustern und Tendenzen unterscheiden? Würden sie Form-Bedeutungs-Paare aufweisen, die bis dato nicht beobachtet wurden?

Auf den ersten Blick mag dies tatsächlich der Fall sein. Bedenken wir beispielsweise, dass Tok Pisin eine komitative und instrumentale Präposition *wantaim* beseitzt, ein Form-Bedeutungs-Paar, das in dieser spezifischen

Zusammensetzung in keiner anderen Kontaktsprache und auch generell in keiner Sprache, die mir bekannt ist, auftritt. Ist eine solche Konstruktion ein Beweis dafür, dass während der Entwicklung der Sprache ein höheres Maß an Kreativität eingesetzt wurde? Handelt es sich bei den Sprechern, die diese und andere Konstruktionen in Tok Pisin geschaffen haben, generell um kreativere Sprecher als Sprecher von Sprachen, die langsamer und mit weniger Einfluss anderer Sprachen entstanden sind?

Diese Frage führt zurück zu einer anderen, abstrakteren Frage: gibt es einen bedeutsamen Unterscheid zwischen den sogenannten Kontaktsprachen und anderen Sprachen? Wenn ja, ist es ein Unterschied in der Struktur, im Gebrauch? Sollte er in der Entwicklung zu finden sein, stellt sich die Frage, was die entscheidenden Faktoren sind, die kreativere Resultate fördern. Handelt es sich dabei um die dringendere Nachfrage und den kürzeren Zeitraum, in denen solche Sprachen entstehen? Oder um eine bestimmte Mischung der Ursprungssprachen? Führt Sprachkontakt zwischen typologisch unterschiedlichen Sprachen zu signifikant unterschiedlichen Resultaten, möglicherweise kreativeren Resultaten, als zwischen Sprachen deren Strukturen sich von Anfang an mehr ähneln?

All diese Fragen beruhen zunächst auf unserem Verständnis von Kreativität. Was bedeutet es, sprachlich kreativ zu werden, und worum handelt es sich bei kreativer Sprache? Ist Kreativität, auf sprachlicher Ebene, mit dem Konzept einer Problemlösung verbunden? Bedeutet Kreativität auf sprachlicher Ebene also, neue Wege zu finden, bekannte Konzepte auszudrücken? Alte Wege zu benutzen, um neue Konzepte auszudrücken? Neue Wege zu finden, neue Konzepte auszudrücken? All diese Dinge? Keines davon?

Selbst wenn wir an diesem Punkt in der Lage wären, diese Frage zu beantworten, blieben wir mit einer anderen zurück: wie, genau, können wir neue Strukturen und Konstruktionen in ein Sprachsystem einführen? Kombinieren wir lediglich bestehende Elemente neu? Führen wir komplett neue Elemente ein? Kombinieren wir bekannte Elemente auf bestimmte Weise, so dass wir Teilaspekte beibehalten, aber auch neue Aspekte einführen? Sind Neologismen unsere Hauptmöglichkeit für Innovation? Bedienen wir uns vor allem bei anderen Sprachen, sofern möglich? Sind wir uns überhaupt bewusst, wieviel unseres Sprachsystems wir (neu) gestalten? In diesem Zusammenhang sind viele Fragen zu stellen. Einige davon werden hoffentlich im Rahmen dieser Arbeit beantwortet.

Zu diesen Fragen gehören fünf Fragenkomplexe, die für diese Arbeit von kritischer Bedeutung sind.

Erstens: was ist Kreativität? Alle übrigen Fragen ruhen auf einem soliden

Verständnis davon, wie die Konzepte der Kreativität und Innovation miteinander verbunden sind, sowohl auf konzeptueller Ebene, als auch auf der Prozessebene innerhalb der Sprache und der Sprachwissenschaft.

Zweitens: gibt es einen signifikanten Unterschied darin, wie frühe Sprecher einer Kontaktsprache Prozesse der Kreativität und Innovation in der Gestaltung der Sprache nutzen, im Vergleich zu Sprechern einer Sprache, in der Kontakt eine geringfügigere Rolle spielt?

Drittens: was genau sind die sprachlichen Prozesse, durch die Sprecher von Tok Pisin neue grammatikalische Strukturen geschaffen haben?

Viertens: sind unübliche Form-Bedeutungs-Paare ein Anzeichen für ungewöhnliche sprachliche Prozesse während der Sprachentwicklung, oder entstehen die Unterschiede durch andere Faktoren? Falls sich die Prozesse nicht unerscheiden sollten, was sind dann die Faktoren, die zu unüblichen Resultaten führen?

Fünftens: wie wird, wenn kreative Lösungen angewandt werden, um kommunikative Probleme in der Entstehung von Kontaktsprachen zu lösen, die Variation zwischen parallel entwickelten Lösungsansätzen verringert und ausgesetzt?

Struktur der Arbeit

Um diese Fragen zu beantworten, ist die vorliegende Arbeit folgendermaßen aufgebaut. **Kapitel 2** legt die für die spätere Analyse erforderlichen theoretischen Hintergründe dar, was verschiedene sprachwissenschaftliche Felder und Theorien angeht. Dabei werden sieben Themenfelder abgedeckt. Das erste dieser Felder legt die Grundzüge der Kontaktlinguistik dar, darunter die Beschreibung der Kontaktsituation, das Entstehen von Kontaktsprachen und ihre Entwicklung sowie ihre Eigenschaften. Desweiteren wird auf das Konzept einer Mischform aus Pidgin- und Kreolsprachen eingegangen, sowie die Bedeutung der Grammatikalisierung für Kontaktsprachen erläutert. Das darauffolgende Themenfeld befasst sich mit Theorien zur linguistischen Kreativität und sprachlicher Innovation. Dabei wird auf unterschiedliche Definitionen von Kreativität u.a. aus der Psychologie eingegangen, um sich einer Begriffsdefinition zu nähern, sowie auf Vorarbeiten im Feld der Sprachwissenschaft. Desweiteren werden die einzelnen Permutationen sprachlicher Innovation dargelegt, und das Verhältnis zwischen sprachlicher Kreativität und sprachlicher Innovation wird festgelegt. Im dritten Themenfeld wird auf Theorien der Evolutionären Linguistik als Zusatzdisziplin eingangen. Grundlegende Konzepte wie Agentivität, die unsichtbare Hand, die Autobahnanalogie sowie

Selektion und Variation sowie ihr Nutzen für die Kontaktlinguistik werden kritisch betrachtet. Das Konzept sprachlicher Komplexität und Adaptivität sowie deren theoretischer Unterbau in Usage-Based-Grammar wird erläutert. Das vierte Themenfeld stellt die Rolle der Größe der Sprechergemeinschaft in der Sprachentwicklung klar, während im fünften Themenfeld die Frage behandelt wird, welcher Teil der Sprechergemeinschaft für die Innovationen zuständig ist. Im vorletzten Themenfeld werden die Begriffe sprachlichen Angebots und sprachlicher Nachfrage eingeführt, bevor im letzten theoretischen Kapitel das Konzept der Grammatikalisierung, insbesondere auf Basis des Reanalyse-Prozesses, näher betrachtet wird.

Kapitel 3 gibt einen Überblick über die sprachexterne Geschichte des Tok Pisin. Dabei wird auf sie soziohistorische Geschichte eingangen, vor deren Hintergrund Tok Pisin und seine Vorgängersprachen entstanden sind. Behandelt werden außerdem die Fragen nach dem Ursprungspunkt der Sprache und den unterschiedlichen Entwicklungsstadien. Der Kontakt mit europäischen Forschern, Händlern und Siedlern auf der einen und den pazifischen Einwohnen auf der anderen Seite wird von seinen Ursprüngen im Walfang über die Handelsoperationen hin zu den Plantagen und der Kolonialisierung bis zu den Weltkriegen und ihrem Nachwirken dargstellt.

Kapitel 4 stellt dagegen die interne, sprachliche Entwicklung des Tok Pisin in den Vordergrund. Aufbauend auf dem vorausgegangenen Kapitel wird die externe mit der internen Sprachgeschichte verknüpft. Sprachinterne Prozesse und Veränderungen die zur Entwicklung von Tok Pisin als eigenständiger Sprache geführt haben, werden dargestellt. Dabei werden die Aspekte zweier der einflussreichsten Modelle der Entstehungsgeschichte, Keesings Ozeanische Substrathypothese (1988) und Mühlhäusler's Entwicklungsstadien (1985) anhand ihrer Voraussetzungen und Implikationen analysiert. Verschiedene Aspekte der Modelle werden evaluiert. Es folgt ein Versuch, die Widersprüche zwischen beiden Modellen auszuräumen und ein gemeinsames Modell zu schaffen.

Kapitel 5 stellt kurz die unterschiedlichen Analysemethoden sowie das zur Verfügung stehende Analysematerial dar. Bei der Hauptquelle für die empirische Analyse handelt es sich um das Z'graggen-Korpus, eine Sammlung von 4.645 Transkriptionen zu Audioaufnahmen, die Pater John Z'graggen seit Anfang der 1970er Jahre vor Ort vollzogen hatte. Größe und andere Eigenschaften des Korpus werden ebenso detailliert und evaluiert wie die notwendigen Arbeiten am Korpus sowie die weitere Textverarbeitung und der Prozess, durch den einzelne sprachliche Proben entnommen wurden.

Kapitel 6 beinhaltet schließlich die detaillierte Analyse von 18 gram-

matikalischen und strukturellen Phänomenen in Tok Pisin und ihrer Entstehungsgeschichte. Dabei wird sowohl auf die Prozesse eingangen, durch die die Konstruktionen enstanden sind als auch auf den Wettbewerb zwischen verschiedenen Strukturen, die dieselbe Funktion erfüllen sowie auf die Faktoren, die den Ausgang dieses Wettbewerbs bestimmen.

Das abschließende **Kapitel 7** diskutiert die Resultate des vorangegangenen Kapitels im Lichte der anfangs aufgeworfenen Fragen; im Zusammenhang mit den verschiedenen Entstehungsmodellen des Tok Pisin; im Zusammenhang mit den theoretischen Ansätzen aus Kapitel 2, sowie im Zusammenhang mit weiteren möglichen Arbeiten. Dabei werden die folgenden Schlüsse gezogen:

1. Tok Pisin zeigt einige ungewöhnliche Strukturen und Konstruktionen auf, allerdings keinerlei ungewöhnliche sprachliche Funktionen

2. Der Ursprung dieser Strukturen und Konstruktionen liegt vor allem in der sprachinternen Entwicklung, nicht in den Ursprungssprachen

3. Äußere Umstände und Timing der Innovationen sind von Relevanz

4. Im Bereich der Syntax lassen sich andere Innovationsmuster feststellen als im Bereich der Morphologie

5. Der Wettbewerb zwischen Formen, die dieselbe Funktion erfüllen, liefert Hinweise auf die Ursprünge dieser Formen

6. Semantische und syntaktische Reanalyse sind von großer Bedeutung für die sprachliche Innovation

7. Innovation und Agentivität der Sprecher unterscheiden sich zwischen sprachlichen Ebenen

8. Die zugrundliegende Entstehung einzelner Konstruktionen spricht gegen Teile der Ozeanischen Substrathypothese

Contents

List of Figures

4

List of Tables

Dedication

Anyone who has ever tried to write a dissertation or a similarly expansive and detailed piece is aware that while the writing itself may be a solitary activity, every single other step in the process is decidedly not. I owe thanks to many who came before me and many who have accompanied me along the way, keeping me inspired, and keeping me sane. Since it would be futile to try and list every linguist and other academic whose works have enabled and influenced mine, I will limit the list below to those I owe some personal gratitude. Even so, this list is not exhaustive and no meaning should be ascribed to the order of appearance.

To my advisors Stephanie Hackert and Hans-Joerg Schmid, for agreeing to read this when I had no idea what it would eventually turn out to be. To Steffi especially, for setting me on the academic path[1] and for opening many doors along the way.

To the PARADISEC project for allowing me access to the data they have collected, without which this work would not have been possible.

To my family. Especially to my parents Christine and Gerhard, for supporting me and encouraging me to live up to my own expectations, even when they might not have matched theirs. To my grandmothers Frieda and Maria, for everything that they've taught me, from the meaning of family down to the cookie recipes (even though I maintain that "half an egg" is not a thing). To my brother Mathias, who in many ways is the better version of me (except for the hair), although he could be even better if he got a dog already. To Desi, as a kid I always thought it'd be great to have a sister, and as it turns out, I was right. But seriously, get started on that dog thing.

To my fellow savages, Alex, Catherine, Diana, Janina and Melanie; to the neighbouring tribesmen and -women Caro, Christoph, Liz and Quirin; to both

---

[1]Sidenote: if one of your professors ever stops you in the hall and asks if you'd like a job after you have done well on an exam, say yes. You'll never know where it leads.

the recently civilized and those still working their way through he PhD jungle, with occasional stops at the Fruppig outpost for f(j)uel.

To the Class of Language, its staff, the associated professors and members, for their feedback and their companionship, with special thanks to Sebastian and the rest of the team for the best and only Heine-approved symposium. Being a small part of everyone else's journey has brought me some valuable perspective (and levelled up my tech skills).

To Christian, for proving that the best of friends may be found in the most unlikely of places. I promise I'll stop grinding for titles now (for a time, at least).

To Kathi, remember that old adage about friends being the family you choose? It's true, and you're the best choice I could have made. Thank you for everything.

To Biene and Markus, for all the coffee and the food and the gossip to accompany it. To walls of code, walls to climb and reasons to stay.

To Nico, Elke, Tichi, Vanessa, Valle and Monika, for drawing me into another world altogether. Maybe I'll finally get to finish Medivh now.

To my fellow trainers Hans, Ferdl, Tobi and Max, as well as all the trainees at Bis behend, to Maxl and all the others beyond. Stay dashing and daring, courageous and caring.

1 Innovation in the emergence of Tok Pisin

1.1 Creativity, innovation and language

The present study analyses the emergence of several grammatical structures in Tok Pisin, the variety of Melanesian Pidgin spoken in Papua New Guinea. In analysing these structures, it shows how speakers of Tok Pisin, at crucial stages of its development, relied on the processes of reanalysis and grammaticalization in order to shape critical structures of the language as it is still spoken today. Rather than relying on borrowing from their various lexifiers, speakers relied on language-internal reanalysis and innovation to solidify an increasingly complex system. In analysing how this process occurred, insight is gained into how languages, or rather, their speakers, cope with functional demand in emerging languages by introducing new elements and recombining existing ones.

The original question sparking this analysis was a seemingly simple one: why is language not weirder? It may be itself a weird question to ask, but consider it for a moment. As a speaker of any language, we are free to invent any word we choose. We are free to change the structure of any sentence. There is no physical limit which keeps us from changing the sounds of a language (aside from limiting them to those our speech apparatus can produce), and there is no particular reason why you could not make up a new tense, with your very own verbal suffix or clitic or free-standing marker to formally express it. So why do we not have as many grammatical systems as we have speakers of a particular language?

The answer - or rather, part of the answer - is, of course, that there are constraints on innovation for any existing language system. The first, major constraint in this system is the one of intelligibility. If I were to change too much about the language I speak, I would no longer be understood by my interlocutors. If I change key parts of a language's structure, my interlocutors

may either fail to grasp my meaning or infer a different meaning, discarding my new, creative way of speaking as an error.

There are situations, however, in which speakers of a language are forced into seeking out creative solutions because either they themselves, or the language system as a whole, may lack a way of expressing a certain concept, be it lexical or grammatical. This is why we have categories such as the Word of the Year, in which we honour new, creative expressions - or recoil in (mock) horror from either them or the concept they describe, depending on our prescriptivist leanings. And there are, of course, changes to the grammatical and syntactical structure of a language over its lifetime. Language as a whole and languages in particular have never been stale, rigid things, but have constantly evolved with their speakers.

We can be creative with language then - to an extent. But to what extent is that? And are there circumstances, or points in the development in the history of a language that force us to be more creative than usual? One possible answer to this question may be a situation in which the emergence and development of a language happens more rapidly and with a more pressing need to develop structure within a language than is normally the case. Such claims have been made for the origins of contact languages such as pidgins and creoles, which are said to develop faster due to their need as a lingua franca among speakers of different languages. One example for these languages is Tok Pisin, as spoken in Papua New Guinea.

Could languages such as Tok Pisin, then, provide clues as to whether our capacity for creativity is unleashed during their emergence? And would that indicate that the structures and patterns such languages develop differ more significantly from typological universals or tendencies? Would they exhibit form-meaning pairings that have previously been unobserved?

At first glance, this may certainly seem to be the case. Consider, for instance, that Tok Pisin has a comitative and instrumental preposition *wantaim*, a form-meaning pairing that occurs in no other contact language, or, for that matter, in this specific form, in no other language period that I am aware of. Is such a structure proof of the superior creativity engaged in the process of creating a language? Are the speakers which originated this and other constructions within Tok Pisin somehow more creative than the speakers of languages which developed more slowly, and with less impact of contact between languages?

This question leads back to a more general one: is there a meaningful difference between the so-called contact languages as a group and other languages? Is this difference one in development, in structure, in usage? If it

should be in development, what are the contributing factors that determine more creative outcomes? Is it just the pressing need and the shorter time period during which such languages emerge? Is it the particular mixture of input languages? Does language contact between languages which are typologically different lead to different outcomes, maybe more creative outcomes than contact between languages whose structure is, from the outset, more similar?

All of these questions rely, first of all, on what we understand by *creativity*. What does it mean to be creative with language, to be creative in language, and what is creative language use? Is creativity, in language, related to the concept of problem solution? Does being creative with language, then, mean finding a new way to express familiar concepts? Using old ways to express new concepts? Finding new ways to express new concepts? All of these? None?

Even if we answer all of these questions, we are left with another: how, exactly, do we introduce new structures and constructions to a language? Do we randomly recombine elements? Do we introduce new elements? Do we recombine elements in a certain way so that we use some of their nature, but also introduce a new aspect? Are neologisms our prime way of innovation? Do we mostly borrow from other language, if possible? Are we even conscious of how much we innovate and create in terms of language structure? There are many questions to ask, and I hope to begin answering some of them.

1.2 Research questions

Drawing on the initial set of questions above, I have determined four sets of questions I will tackle in this thesis.

First, what is creativity? All remaining questions rely on a solid understanding of how the concepts of creativity and innovation link up and relate to each other on both a conceptual level and a process level within linguistics.

Secondly, is there a significant difference in how speakers of emerging contact languages use creative and innovative processes to shape their language when compared to speakers of languages in which contact does not play a large role?

Thirdly, what are the concrete linguistic processes by which speakers of Tok Pisin have introduced new grammatical structures into their language?

Fourthly, are unusual form-meaning pairings indicative of unusual processes in language development, or does the disparity come from other sources? If so, and the processes do not differ, what are the factors which determine typologically unusual outcomes?

Finally, if creative solutions are applied to communicative problems in emerging contact languages, how is the variation between parallel innovations levelled?

1.3 Overview

In order to answer these questions, this study will proceed according to the following structure: **Chapter 2** provides the theoretical background on various linguistic fields and theories which inform the analysis in chapter 5. It covers five major topical groups. The first includes the bases for contact linguistics as they pertain to the description of the contact situation, the emergence of contact languages and their development and general features. Also included are theories of evolutionary linguistics as a supplemental discipline in regards to the emergence of specific features, including variation and competition between features serving the same grammatical function. The second group is concerned with the phenomena of linguistic creativity and linguistic innovation. Innovation and creativity as well as their cognitive bases, linguistic expressions, their potential and their limits in regards to linguistics and the question of who innovates, who selects and who regulates are detailed from various approaches. It also introduces the question of whether the economic concept of supply and demand can be transferred to (contact) linguistics, and how it could be applied to concrete contact situations. The third group is concerned with the effects of the size of the speech community on linguistic development, as well as the definition and effect of linguistic complexity on a language system. It also covers the idea of language as an adaptive, complex system and the requirements of such as model to be based in usage based grammar rather. Finally, the fourth group of theories relates to the process of grammaticalization, its' subprocesses, the notion of grammaticalization sources and paths, as well as the notion of contact-induced grammaticalization.

Chapter 3 details the external history of Tok Pisin. It gives a brief overview of the sociohistorical and historical background against which Tok Pisin and its precursors emerged. It covers the questions as to where to set the beginning of Tok Pisin and goes on to explore the various stages of contact between European explorers, traders and settlers and the Pacific Islanders which were native to the region of Melanesia from early whaling and trading operations through various colonizations, plantations, wars and independence.

Chapter 4 details the internal, i.e. linguistic development of Tok Pisin. Building on chapter 3, it aligns the external history of the speech community

with the language-internal processes and changes which led to the emergence of Tok Pisin as an independent language. It covers the aspects of the two main competing models, which are Keesing's Oceanic substrate hypothesis and Mühlhäusler's developmental stages, as well as their presuppositions and the implications each of the models has for the phenomena analysed in later chapters. Various aspects of each model are evaluated both on their own merit as well as in a broader scope that takes into account recent scholarly work on Tok Pisin and the relevant features. The models are also evaluated against the available sociohistorical information, such as the size and structure of the speech community at various stages of development. Finally, the models will be separated into individual claims and implications to gauge whether a combined model would be possible or whether they inherently preclude each other. At the end of this process, a unified model will emerge, or one of the models will be discarded in favour of the other. A unified model may also incorporate more aspects of one original model than of the other.

Chapter 5 details the methods and material employed in the analysis of the various grammatical phenomena which form the core of this thesis. The main source of original language data, the Z'graggen corpus, which is based on some 4,645 texts in Tok Pisin recorded by Fr. John Z'graggen from the early 1970s on, is described in detail as to its content and structure. Size, token number, type number and type/token ratio of the corpus are detailed and evaluated, as well as the text length of the individual texts within the corpus.

The chapter also includes detailed information on how the corpus was processed in order for it to be analyzable for this project, including the deduplication of the texts, text processing and the sampling for the individual phenomena surveyed. In an effort to be as open and replicable as possible, this chapter includes the commented R code for all steps in said process, as well as a detailed step-by-step explanation of the annotation process in Partitur, where it was carried out.

Next, it describes how and why the specific grammatical phenomena analyzed in the following chapter were selected and how the initial sample size for each of these phenomena was determined. Finally, it lays out the steps in which the items and patterns of interest would be analyzed in the following chapter.

Chapter 6 contains the detailed analysis of all grammatical structures and constructions found to be a of primary interest to the questions detailed above. It establishes the processes by which they originated and diffused in the speech community and identifies competition between variants as well

as factors which may have led to the selection of one variant over another.

Chapter 7 discusses the results obtained in the preceding chapter in light of the questions raised in the beginning, their relation to various theoretical concepts detailed in chapter 2, including the competing models of the stabilization of Tok Pisin, and the notion of competition and selection as well as other notions from Evolutionary Linguistics. It also relates other patterns found across the structures and their origins and lists the conclusions which may be drawn from them.

2 Theoretical Foundations

2.1 Pidgin and creole formation

2.1.1 Definition of pidgin and creoles

This chapter will provide the theoretical background informing the analysis performed in chapter 6 below. The first linguistic field relevant to said analysis is the one of language contact, more specifically the origin and development of pidgins and creoles. Linguists agree, for the most part, that these are two possible outcomes of of two or more languages - or rather, their speakers - coming into contact. There is less agreement on their specific definition and shared characteristics, however (see among others, Bickerton (1977, 49), Crystal (1987, 334), McWhorter (1995, 240), Mühlhäusler (1997, 6), Winford (2009, 18f.) and Yule (1996, 233-234) for definitions of pidgins).

Velupillai (2015, 15) gives perhaps one of the most accessible initial definitions and this work will largely follow her: "[a] pidgin is, very simplified, a language that emerges when groups of people are in close and repeated contact, and need to communicate with each other but have no language in common." She also points out (2015, 16) that there are three crucial elements to this definition:

> First of all the contact needs to be repeated or extended. A spontaneous communicative bridge between, for example, a Swedish tourist who only speaks Swedish and an Italian fruit vendor who only speaks Italian, involving many gestures and efforts to understand each other, is not a pidgin but rather a jargon [...].
>
> Second, in a contact situation that gives rise to a pidgin, the languages of each party are typically not mutually intelligible. Danish and Swedish, for example, are two very closely related languages that to a great extent are mutually intelligible. A Danish tourist speaking a kind of temporarily stilted and "Swedified" Danish when negotiating with a Swedish fruit vendor is not speaking a pidgin. [...]

> Third, a pidgin is not simply an imperfectly learned second language. For example, speakers that are trying to learn a new language will be at various levels of competence in it while they learn it and might simplify the target language in various ways.Or immigrants may be imperfectly competent in the language of their host country. In neither of these situations are the speakers using a pidgin [...].

Creoles are also difficult to define, as, among others, Winford (2009, 22) and DeCamp (1977, 3-4) admit. In general, there are three possible approaches (see, for instance, Hickey (2014, 84), Holm (1994, 372) and Mühlhäusler (1997, 7)): either by their (putative origins), their communicative function or their structural characteristics. In the first conception, creoles are merely seen as pidgins which have undergone an elaboration process through the adoption as native languages. While that process can lead to creolization, and usually has heavy implications for the further structural development of the pidgin in question, not every creole has its origin in pidgins. Neither is it possible to define a creole as a pidgin which has adopted additional communicative functions, i.e. a pidgin used in additional contexts beyond trade, negotiation or whatever its original function may have been. Finally, the definition of creoles based on structural characteristics is also difficult at best, as there are no structures and features common to all creoles. For this reason,the definition of a creole on structural ground remains disputed at best. A general definition will therefore have to be reduced to fit most (if not all) languages deemed to be creoles. As such, this work once again largely follows Velupillai (2015, 43) in defining it as "a natural language spoken as a mother tongue by an entire community that arose due to situations of intense contact" and "[a] fullfledged languag[e] on par with any other natural language in the world, [...] capable of fulfilling any linguistic need of the relevant speech community." It is the first part of the definition that is especially crucial, defining creoles by their origin or their socio-historic history. As Wardhaugh and Fuller (2015, 128) point out, "it is only because we know the origins of creoles that we know they are creoles". This foreshadows one of my central points: creoles are not special in the linguistic processes that they are subject to, but in the circumstances of their origin.

2.1.2 Origin of creoles

Unfortunately, but not surprisingly, theories of creole formation are at least as disputed as the creole definition itself. Winford (2009, 329), however, quotes a "recent compromise" which claims that "creole formation involves varying degrees of input from both superstrate and substrate and is guided by principles that regulate all cases of language contact". While such a compromise may seem a bit abstract at first, it precludes extreme positions such as creole exceptionism and exclusively universal-based accounts.

One of the main controversies in the emergence of creoles has been the question of whether they have emerged abruptly or gradually. The Gradualist Model as detailed in Arends and Bruyn (1994) claims creole formation to not be an abrupt or instantaneous nativization of a pidgin, but instead a gradual process extending over generations of speakers. They cite the elaboration of pidgins into full-blown languages such as modern Tok Pisin as evidence for such a hypothesis, along with other developments such as the split within the copula system in Sranan, which began in the first half of the 19th century and thus 150 years after creolization had set in. In the case of the latter, the development is linked to demographic changes (the balance between locally-born and African-born blacks shifting towards the former, see Arends and Bruyn (1994, 114).

The underlying question in this regard is whether pidginization and creolization are processes which are distinguished from 'ordinary' language change only in their speed, or whether they are qualitatively distinct. Complicating this question is the fact that not all creoles arose under the same circumstances. This, however, can also be a boon in categorizing them. One such possibility is given by Velupillai (2015, 48-52), based on the socio-historic origins of the individual creoles. The main line she draws is between exogenous creoles, which are "those which arose in a setting where none of the groups involved were indigenous to the area".

As she puts it, the crucial factor for exogenous creoles is that both the speakers of the superstrate **and** the speakers of the substrate languages were immigrants to the area in which the creole emerged. Their social status and numerical ratios, however, were not equivalent, with the labour force usually being in the numerical majority, but of lower socioeconomic status. In addition, substrate contact was minimized after resettlement of the labour force. Subtypes of this category include plantation creoles and maroon creoles. The first refers to creoles arising on plantations, whereas the second describes creoles arising between groups of slaves who had escaped plantations and formed

communities on their own, but still lacked a common language.

Opposite exogenous creoles in this model are endogenous creoles, which "typically developed through contact between immigrant settlers, usually engaged in systematic trade, and the indigenous population of the areas". The latter, however, is very close to the definition or pidgins.

This raises another dispute in regard to the origin of creoles: their connection and relation to pidgins. In many early accounts of language contact Schuchardt (1914), Jespersen (1922), Bloomfield (1933) as well as more recent accounts such as McWhorter (2018), creoles are said to derive from pidgins, whereas other accounts frequently dispute such a connection. Mufwene (2008, 75-78), for instance, dedicates a whole subsection to "Why creoles have not developed from pidgins", while McWhorter argues that creoles betray structurally betray their origins as pidgins, citing, among others, Tok Pisin as one of the demonstration cases (McWhorter 2018, 14).

Mufwene's argument is that given that the prototypical creole has evolved in plantation settings, interaction between the Europeans and non-Europeans, the latter forming the majority, would have been conducted mostly in the European language during the initial homestead phase. He bases this assumption on the idea that most non-European languages would not have had a sufficient number of native speakers to be immediately useful as a lingua franca between sizeable groups. Therefore, and due to the higher social prestige of the European languages, "the earliest vernaculars commonly spoken and appropriated by non-Europeans were approximations of the European colonial languages". Creole children growing up in the same homestead together would not have exhibited great linguistic variety, and modelled their speech not on their parents, who were at work, but the community at large.

Eventually, fuelled by the majority of the population being non-native speakers of the European language, as well as by rapid population turnovers and demographic growth, "the increasing demographic dominance of non-native speakers among non-Europeans communicating primarily among themselves in the new vernacular also favoured a greater role of non-European substrate influence". It is during the plantation phase, not the homestead phase, that "the local European language gradually evolved into a different variety among the non-Europeans" (2008, 76f.).

The circumstances in origins for creoles, therefore, differed greatly from that of pidgins, in which contact was decidedly limited. In plantation settings, there was no initial sporadic contact, but rather immediate, long-lasting contact. Why then, would we see structural similarities between pidgins and creoles? According to Mufwene (2008, 78), such similarities "reflect the fact that

they were developed largely by linguistic adults interacting regularly among themselves, using materials from typologically related European and/or substrate languages to meet diverse and complex communicative needs, and thus needing complex grammatical structures".

If Mufwene forms one end of the debate on whether creoles (necessarily) derive from pidgins by answering that question with a no, the opposite end of the scale is personalized by John McWhorther, who enthusiastically answers it with yes. Opposed to what he calls the Uniformitarian Hypothesis, an umbrella term which he assigns to theories that creoles do not derive from pidgins, McWhorter has long been vocal in The Creole Debate, which also happens to be the title of his latest book. Unlike Mufwene, he "unhesitantly embraces the the pidgin-creole life cycle as a defining and predominant scenario of creole language genesis" (McWhorter 2018, 9), arguing that creoles result from expanding and nativizing pidgin languages when they come to be used by a first generation of speakers (a stance once referred to as Creole Exceptionalism). Creoles, he claims in his Creole Prototype, show undeniable traces of their origin in pidings, namely, the absence of compositional derivation, lexical tone and productively used inflections (McWhorter 2018, 21). These, he says, are only shared in their entirety by creole languages (though there are plenty of languages which show one or two of them), and only because they have a common origin in pidgins.

In the end, then, the consensus of where pidgins and creoles originate is a very thin one: in language contact. Whether to draw some additional dividing line among the origin contact languages is up to the individual linguist. For the remainder of this work, I will operate under the assumption that the initial contact setting is insufficient to determine what form - pidgin, creole, or otherwise - the emerging language will take. As we are about to see, the forms are not easily distinguished in any case.

2.1.3 Common structural features of pidgin and creoles

While chapter 2.1.1 above has pointed out that there is no set of linguistic features that all pidgins or creoles share, there are some features that are common to many of them. Winford (2009, 270) claims that pidgins, for instance, often exhibit the following characteristics:

1. absence of morphological apparatus such as affixation and inflection
2. absence of other functional categories such as tense and aspect

3. minimal inventory of function morphemes (articles, prepositions, etc.)
4. restricted number of question words and pronouns
5. use of only one universal negative marker
6. analytic structures, constituent function primarily determined by word order
7. reduced number of sentence patterns (e.g. no inversion for questions)
8. lack of derivational depth, no subordinate or embedded clauses
9. generic lexical entries with high ambiguity and polysemy
10. expansion of lexicon through compounding, metaphorical extension and borrowing
11. vocabulary taken primarily from one contact language
12. reduced inventory of phonemes and phonological contrasts

In the same vein, there are a number of features that many creoles are said to share, among them the following (2009, 319f.):

1. **phonological:** tendency to keep vowels/consonants common to source languages; substrate influence on phonotactics (e.g. CVCV syllable pattern); cluster simplification
2. **lexicon:** bulk of input from superstrate, retentions from substrate, internal innovations; semantic reanalysis (semantic+categorial change) under substrate influence; compounding modelled on both superstrate and substrate patterns
3. **morphology and morphosyntax:** general lack of inflectional morphology; case and gender distinctions reduced or eliminate in pronominal system while number is preserved, lack of copulas in ascriptive-type predication
4. **syntax:** elaborate syntactic systems; identificational (contrastive) focus constructions and varieties of serial verb constructions (fulfilling functions associated with categories such as prepositions and complementizers in lexifier language)

While the lists above contain items such as 'reduced inventory' and 'lack of derivational depth', it is important to point out the following: in discussing the circumstances of their origin, creoles have, in the past, been described as 'failed' or 'broken' transmission of both a target language (such as French, English or German in overseas colonies) and an ancestral language (the languages of the colony's respective native inhabitants). Basically, the argument has been that socio-economic or historic factors made it impossible for speakers to fully acquire the target language, leaving them only with bits and pieces.

Baker (1994, 65) summarizes these views as the superstratist, universalist and substratist positions, each of which assumes a measure of failed acquisition.

The implication often has been that these languages are somehow 'lesser' or incapable in some way, with either universal language features or resources from the ancestral language being employed to 'repair' gaps in the transmission and missing language features in the 'broken' version of the target language that had been acquired. Creole speakers were thus seen as having been unsuccessful both in acquiring the imported language and at the same time, in retaining their ancestral language. Fortunately, this view is no longer prevalent in the current literature due to solid empirical work such as that by Philip Baker (1994). As he and others have argued, creoles and pidgins "[are] not and [were] not the result of unsuccessful attempts by immigrants", but rather solutions in "overcom[ing] a problem of communication", a task for which they drew "on the full range of resources available to them" (1994, 83). In other words: the full array of linguistic features in creoles and pidgins show that they not only drew on existing languages, but rather that they are "innovations which took place during the recorded history" (ibidem) of their creation. As thus, speakers should be credited for solving communicative problems with available resources rather than failing some perceived standard of language acquisition and using material from other languages as a stop-gap measure.

Back to the listed structures, similar lists can be found in Velupillai (2015, 53-55, with the caveat of merely being "the features that are commonly brought up in the discussion of the possible uniqueness of creole languages"), Bartens (2013a, 83-146), Markey (1982), Taylor (1971), Baker and Huber (2001) and many others. Anyone willing to hunt down the various lists will quickly notice that few of them match closely, strengthening the argument against defining creoles on the basis of common features. There is an argument to be made, I suppose, for ascribing to creoles the notion of a Wittgensteinian family resemblance, in that they need not share the entirety of features on the list, but frequently show some combination of them. The question, then, would become whether such family resemblances among their features cannot as easily be established between any number of arbitrarily chosen languages.

2.1.4 The notion of elaborated pidgins or pidgincreoles

Early Tok Pisin and its predecessors in the developmental chain show, as we will see below, many of the features listed for pidgins above. However, later Tok Pisin and especially modern day Tok Pisin no longer do. If Tok Pisin

resulted from a pidgin, but pidgins do not necessarily evolve into creoles as described above, that calls for a category of 'advanced' or 'elaborated' or 'extended' pidgins.

According to Winford, Tok Pisin falls exactly into this category of 'elaborated or extended' pidgins (Winford 2009, 20f.). Velupillai (2015, 20f.), who calls this category "pidgincreoles", agrees, specifically quoting Tok Pisin as an example of this type of contact language, naming it along with Nigerian Pidgin. For them, these are characterized by more complex lexical, grammatical and structural features than the 'pure' pidgins described above. As such, they are structurally nearly or completely indistinguishable from creoles. As their name suggests, they are former pidgins who have undergone elaboration in structure, are capable of being employed in a greater variety of situations and show an increase in structural norms (Velupillai 2015, 21), which Winford argues happens due to interplay between internally driven development and substrate influence (2009, 288-297). As the full lexical and grammatical input of the lexifier language is unavailable to the learners due to the socioeconomic situation under which the elaboration happens, speakers draw increasingly on their own L1 knowledge and on creative internal restructuring to expand the resources of the pidgin. In addition, they are mother languages to some speakers, which pidgins, they argue, cannot be.

In general, this category offers one specific advantage over earlier models: it allows for processes of elaboration, extension, innovation, complexification and diffusion to affect any type of contact variety, i.e. all contact varieties and languages may be subject to the same changes, and it is only the sociohistoric circumstances under which they arise that lead to different output. These circumstances determine the input, both intra- and extralinguistically, which, in turn, determines the outcome.

In this regard, as we will see below, it is the "creative internal restructuring", as Winford calls it, which leads to many of the more uncommon structural features of Tok Pisin. It is important to note, however, that this creative restructuring does not involve any processes that do not happen in most other languages of the world. It is the circumstances they happen in and the material that is available to be processed that differs in situations of extended language contact. Whether or not this means that speakers that originated these uncommon features were more creative and innovative requires a closer look at the notion of creativity and innovation, which will be the focus of the next chapter.

2.2 Linguistic creativity and innovation

Given that the main interest of the present study lies in the innovation of linguistic constructions, and the creative use of available material to do so, it needs to first be determined what exactly is to be understood by linguistic creativity and linguistic innovation.

In trying to establish a working definition of what linguistic creativity actually entails, it may be beneficial to investigate definitions of creativity in other scientific fields, such as psychology. Franken (1998, 353) provides the following three reasons for people to engage in creative acts:

1. The need for novel, varied, and complex stimulation. One way to meet this need is to create or find new things that will stimulate our senses (for instance, new recipes, new art, or new cars) or challenge our intellect (for instance, books, computers or movies). Berlyne (1960) has attributed the creation or appreciation of beauty to this need.

2. The need to communicate ideas and values. Concerned that children are dying of starvation, a photographer triggers our compassions with a picture of an emaciated child. A politician wanting to make a difference writes a book to challenge our beliefs and stimulate us to action.

3. The need to solve problems. As we encounter new diseases or our business begins to fail, we search for answers that give us hope.

Let us take a look at how these three reasons influence linguistic creativity in particular, by going through them one at a time, starting with the last. At a cursory glance, it might seem like the third reason is the most compelling when it comes to linguistic creativity (and the second when it comes to "normal" use of language). As a group of language users, we may encounter – or invent – new objects, new events, new experiences, which, in their novelty, surpass our capacity to express them within the set of lexemes that are already available. While we may simply extend the semantic extension of lexemes which already exist, we may also invent entirely new lexemes or recombine existing lexemes or morphemes. Consider, for example, the lexeme *smog*, recombined from parts of smoke and fog through blending, or the introduction of trade names such as in the classic example of creatio ex nihilo: that of George Eastman's Kodak in 1888. It therefore seems natural to assume that the third reason is the driving force behind most instances of linguistic creativity.

However, the other two reasons should not be disregarded. The second

reason Franken cites – that of the need to communicate ideas and values - is, in fact, particularly important to human language in general and linguistic creativity in particular. The most basic function of language is, after all, to provide a tool for human communication. Once this need passes beyond the immediate necessities of human life and no longer is limited to referential functions, but includes the emotive, conative and phatic language functions, the opportunities for creative use of language start to emerge. Anyone involved in politics or other areas of society which rely heavily on public relations is bound to be aware of this fact: there is no need for an actual novelty of ideas, objects, beliefs, products or anything else in order to come up with new words for it. Marketing strategists, as well as political advisors, have known about the power of rebranding – essentially, using a new term for an existing brand or other idea, event or person to be designated and thus mixing the second and third motivation – for decades. Whether the reason behind the rebranding is to rid a company of a negative public image, such as the Blackwater mercenary company frequently changing its name (Ukman, 2011), or governments trying to avoid the word war in relation to the armed conflicts they are engaged in (Young, 2014), the goal is the same: to associate the new identity of the company, political party or product with the desired new traits instead of the undesirable old traits. Naturally, the process works just as well the opposite way: find a pejorative term to describe your political opponent's agenda and hope it sticks with the general public. The Republican Party in the United States, for instance, has used Obamacare as a derogatory term for the Affordable Care Act for years (Cillizza, 2012). Whether through the creative combination of existing lexemes or through the coinage of an entirely new term, political and public relation strategists are always on the lookout for salient terms to describe products, ideas, events and opponents. The communication of ideas and values is the underlying motivator of such branding and rebranding. Franken's second reason for creativity therefore influences linguistic creativity as well. In the instance of language, we may question whether the second and third motivation can be separated, given that the essential use of language is communication, but they both definitely play a role.

The influence of Franken's first reason – the need for novel, varied, and complex stimulation – is a bit more complex. One might question how exactly the products of linguistic creativity stimulates the senses. And certainly, a new word or a new way of expressing a familiar concept will, to the vast majority of people, may seem far less exciting and stimulating than trying an unfamiliar dish or driving a new car. Part of Franken's first reason still remains, however.

The process of engaging in linguistic creativity matches his idea of challenging our intellect. In trying to come up with creative ways to communicate, we have to fall back on our knowledge about language and actively employ it in the search for possible expressions – while still being mindful about the various rules that limit these expressions. This challenge becomes most evident when we allow for our concept of linguistic creativity to move beyond the creation of single words or new structures within an existing language and apply to the construction of entirely new languages, the so-called constructed languages or conlangs. J.R.R. Tolkien, in his 1931 lecture A Secret Vice, as follows:

> This idea of using the linguistic faculty for amusement is however deeply interesting to me. [...] The instinct for "linguistic invention" the fitting of notion to oral symbol and pleasure in contemplating the new relation established, is rational [...] In these invented languages the pleasure is more keen than it can be even in learning a new language [...] because more personal and fresh, more open to experiment of trial and error. And it is capable of developing into an art, with refinement of the construction of the symbol, and with greater nicety in the choice of notional range. (Tolkien 1931, 206)

Tolkien had long thought himself alone in the endeavour of making up new languages purely for his own amusement until a chance encounter with a man who, while standing next to him, "said suddenly in a dreamy voice: 'Yes, I think I shall express the accusative case by a prefix!' A memorable remark!" (Tolkien 1931, 199). He found that in creating a language from more or less nothing, "[there are] no base considerations of the 'practical', the easiest for the 'modern mind', or for the million - only a question of taste, a satisfaction of a personal pleasure, a private sense of fitness". Tolkien's description of the inherent pleasure in creating a new language, with which the growing conlanger community of today would probably agree, matches Franken's first reason for the motivation of creativity perfectly: an intellectual challenge, a task that requires thought and knowledge akin to being moved to deep thoughts by a good book or, even, the creation of a piece of art. In summary, all the reasons for creativity that Franken cites as its basic motivations apply to linguistic creativity as well, though their impact may differ.

2.2.1 Definitions of creativity in psychology

While we now know that Franken's three reasons can also be used to explain why people feel the need to be linguistically creative, they still do not explain what creativity is in and of itself and how we may define it. In pursuing this

question, we turn first to traditional psychology, a discipline in which "the [...] definition of creativity includes two parts: originality and functionality" (Kersting 2003). Originality is fairly straightforward: we would not describe something as creative if we have encountered it several times already. We can assume that this part of the definition applies to the subset of linguistic creativity as well. Functionality, on the other hand, is less obvious, but just as important. In psychology, this part of the definition is sometimes divided into quality and relevance, i.e. the creative idea has to be "of high quality [and] must also be appropriate to the task at hand or some redefinition of that task" (Kaufman and Sternberg 2010, xiii). It is not sufficient to be merely original, then. The original thought must also have an actual use. After all, anyone could come up with an original object or a new lexeme. If I were to make up a lexeme like *kwenthaunian* right now, it would almost certainly pass any test on originality. Without an actual use for this new lexeme, however, it is doubtful that it would ever enter into general usage within the English or any other language. In fact, it would have as much use as if I painted a banana blue, put it on a stick and declared it a creative innovation. It might be amusing for a couple seconds, but it really serves no actual function in a broader context.

Just like the motivations for general creativity, then, it may seem that at least two of the general characteristics of linguistic creativity and general creativity also match: originality and usefulness must be present. The notion of "high quality", however, does not apply in linguistics. There is no way to judge a linguistic feature to be of high or low quality on its own merit. There is no phoneme, for instance, that is inherently of higher quality than others. We could establish a criterion of high quality on the basis of how well the newly created item fits into the existing language system it is introduced to, but this would just be the criterion of originality in disguise. We will return to the notion of diffusion below. For now, it will be sufficient to establish originality and usefulness as two criteria for linguistic creativity.

Another interesting notion from psychology is that of the so-called little c and big C. The first, according to Kaufman and Beghetto (2009), is thought to relate to "those creative activities in which the average person may participate each day (e.g. creatively arranging family photons in a scrapbook; combining leftover Italian and Chinese food to make a tasty, new fusion of the two cuisines; or coming up with a creative solution to a complex scheduling at work)". Big C, on the other hand, denotes true creative greatness, being concerned with "creative works [that] may last forever" and which have forever changed their respective fields in some way (2009).

This notion would translate well into existing linguistic frameworks. After

all, everyone might engage in linguistic creativity in a small manner on a day-to-day basis, while few of us get the opportunity to shape a language as a whole. As Ronald Carter (2004, 13) put it, "[l]inguistic creativity is not simply a property of exceptional people but an exceptional property of all people". We should therefore not associate creativity and innovation with very few, exceptional users of a language rather than the speech community as a whole, which finds creative solutions to crucial communication problems. As Carter's quote illustrates, the creative and innovative use of language is something that all speakers and writers engage in to some extent. The introduction of new linguistic items and constructions through the processes of innovation and creativity is a fundamental part of language change, just as the reduction and elimination of other elements, which does not rely on especially uncreative individuals, either.

2.2.2 Linguistic factors of creativity

If every single one of us is capable of being creative in our language use as established above, does that mean that any time we use language, we are creative? This is where we get into the actual linguistic factors in linguistic creativity, and to previous definitions of linguistic creativity in the literature. Zawada (2009, 12) summarizes them as follows:

> [V]arious linguistic researchers have various conceptions of what the term linguistic creativity actually refers to, i.e. there is no general agreement on the nature of this creativity. For some, there is a focus on the combinations of known elements in a new sentence (which we can call generative creativity), whilst for others, there is a focus on the lexical aspects of creativity by which new concepts are created and named, for example through metaphor (which we can call lexical creativity). As Hudson (2000: 10) points out, the creativity of language is due to the openness of language (i.e. the ability to create new lexical items, i.e. lexical creativity), as well as to recursion (i.e. the reuse of syntactic patterns within larger patterns, as well as the re-use of syntactic patterns for new instantiations, i.e. generative creativity).

Here, we see two very different approaches to linguistic creativity. Whereas one - lexical creativity - centres around the introduction of new concepts into the language system, the other - generative creativity - is concerned with the recombination of existing elements. Any general theory of linguistic creativity, however, would have to account for both of these levels: introduction of new elements **and** recombination of existing elements, no matter whether these

elements are lexical or grammatical in nature.

This might, in fact, be a good point at which to establish what prerequisites a theory of linguistic creativity would have to fulfil. I would argue for the following:

1. not language-specific, but valid across all languages
2. not speaker-specific
3. has to address whether creativity requires conscious innovation
4. has to answer the question of what creativity is in the first place
5. has to answer the question of whether all utterances are linguistically creative
6. has to address whether/how linguistic creativity can be empirically observed
7. has to address structural and conceptual levels as well as lexical and grammatical levels
8. has to address the issue of opposition between generative creativity and lexical creativity
9. has to address the levels of language that can be creatively used/changed

Zawada's (2009, 77) final definition covers many of those points:

> (Linguistic) Creativity is an essential and pervasive, but multi-dimensional characteristic of all human beings (irrespective of age, education, intelligence, social status or (non)-artistic bent). Linguistic creativity is primarily the on-line activity of making new meaning by a speaker (in the broadest sense of the user of language in all forms and in all mediums), and the re-creation and re-interpretation of meaning(s) by a receiver. Linguistic creativity is secondarily observable as a feature or product in a language. Linguistic creativity is a graded phenomenon ranging from the more conventional and predictable to the less conventional and unpredictable, and it is manifested in all domains of language (lexis, grammar, text and discourse, as well as medium), the results of which may or may not become conventionalised and therefore entrenched in a particular language.

In particular, this definition covers the first two prerequisites by defining creativity as an "essential and pervasive [...] characteristic of all human beings". By referring to it as an "online activity", it suggests that it is a conscious process. It defines creativity as "making new meaning". By acknowledging that it is a graded phenomenon, it acknowledges that individual utterances can be more or less creative. It puts the onus of empirical observation on features and product rather than the process itself, which remains unobservable. All

of these points are therefore covered sufficiently.

However, it is the last three that are still lacking. Zawada attempts to solve the distinction between structural and conceptual creativity by placing them on a continuum, as seen in figure 2.1[1] (adapted from Zawada 2009, 79). The position on the continuum is determined by "the degree to which there is

structural productivity

conceptual creativity

greater reliance
on existing base,
outcome more predictable
e.g. WFRs like derivation
and endocentric compounding

less reliance
on existing base,
outcome less
predictable
e.g. metaphor,
metonymy and word-
creation strategies

Figure 2.1: Continuum of creativity

a precedent in the linguistic repertoire of the language on which to base the creativity", with structural productivity being "the predictable and productive re-creation of many instantiations based on the same structural pattern or rule", whereas "pure semantic innovation, here called conceptual creativity (i.e. the creation of new and unpredictable concepts)" would be located at the other end. The problem with such a model of creativity is that it does not account for simultaneous creativity on various linguistic levels, or for the innovation of new grammatical elements, rather than concepts. Creativity, however, has more permutations than that.

2.2.3 Permutations of innovation

Language inherently consists of two levels, form and function. It is possible to be creative on one such level, but not on the other, to be creative on both levels, and to be creative on neither. Instead of a two-sided definition of linguistic creativity that only includes lexical and generative creativity, we have to consider a much more complex picture with three levels of creativity.

The first distinction on creative levels to be made is the one between form and function. Any linguistic item or pattern can be innovative on neither, both or one of these levels. There is a further distinction made between syntactic function and lexical function on the function level.

---

[1]The abbreviation WFR in the figure stands for "word formation rule".

Conceptual innovation: the semantic properties of the linguistic item or pattern fulfil the criterion of originality, i.e. a meaning is expressed through the use of the linguistic item or pattern that could not previously be expressed in common usage of the language in question by using a single item or construction. An example would be the addition of *cyberspace* to the English lexicon, introducing a single lexeme to describe a concept that would otherwise require circumlocution.

Morphophonemic innovation: the morphophonemic properties of the linguistic item or pattern fulfill the criterion of originality, i.e. the morphemes and/or phonemes contained within the item or pattern have not yet been present in common usage of the language in question either as individual morphemes/phonemes or in combination with each other. An example would be the German *Handy*.

Morphosyntactic innovation: the morphosyntactic properties of the linguistic item or pattern fulfil the criterion of originality, i.e. the use of the item or pattern encodes morphosyntactic information that could not previously be expressed in common usage of the language in question by using a single item or construction. An example would be the adjective marker *fellow* in Pacific Jargon English, which would eventually become the adjective suffix *-pela* in Tok Pisin.

These three distinct levels indicate that there are eight permutations of linguistic innovation, as illustrated in table 2.1 below:

Concept	Morphophonemic	Morphosyntactic	Example
innovative	innovative	innovative	conlangs
innovative	innovative	not innovative	neologism
innovative	not innovative	innovative	???
innovative	not innovative	not innovative	meaning expansion
not innovative	innovative	innovative	???
not innovative	innovative	not innovative	rebranding
not innovative	not innovative	innovative	syntactic reanalysis
not innovative	not innovative	not innovative	idiom

Table 2.1: Permutations of innovation

The first permutation, in which all three levels are innovative, rarely happens in existing languages, since it requires a new concept, a new form, and a new syntactic function all at once. This would mean that none of the three levels could motivate the other. It is therefore relegated to highly agentive linguistic innovation, such as in the construction of artificial languages ('con-

langs').

In the second permutation, the conceptual and morphophonemic levels are innovative, whereas the morphosyntactic level is not. This would describe the traditional notion of neologisms, as they encode new concepts with new forms.

The third permutation, in which the conceptual and morphosyntactical level are innovative, is something of an issue in this model. It does not seem to occur.

The fourth permutation, in which the conceptual level is innovative, but none of the other two are, would describe meaning expansion, i.e. a change purely on the semantic level.

The fifth permutation, in which the conceptual level is not innovative, but the other two are, describes a process in which a morphosyntactic marker would be directly innovated in a new form. This, too, does not seem to happen, given that syntactical functions are not innovated with a lot of agency.

The sixth permutation describes a process in which conceptual and morphosyntactic levels are unaffected, whereas the form changes. While rare in language change, it's possible to change the names of brands, for instance, in rebranding exercises.

The seventh permutation describes reanalysis: an existing item gains a new morphosyntactic function. The example *-pela* from above works for this permutation.

The eighth permutation, in which no innovative level is engaged, can best be exemplified by quoting an idiom, i.e. an utterance that has been used in that form and function numerous times before.

The critically engaged reader will have noticed that rather than referring to linguistic creativity in these permutations, I have used the term innovation. This brings us to an important distinction between just these two phenomena.

2.2.4 Linguistic creativity vs. linguistic innovation

We arrive at the question, then, of how to differentiate linguistic creativity from linguistic innovation. One characteristic is that of agency. As Mufwene (2008) notes:

> [. . .] speakers are the "unwitting agents of [language] change." They bring this about through repetitions of their own or other speakers' innovations or through repetitions of the recent modifications of older patterns in their communicative acts.

These changes, however, do not have to happen consciously, as we will see below, as well. The relationship between linguistic creativity and linguistic innovation is one of the former being an integral part of the latter. Linguistic innovation is one part of language change, the introduction of new linguistic elements to a linguistic system. The same holds true for the distinction between linguistic creativity and linguistic innovation. The part of innovation which happens on a conscious level, i.e. with the agency of the speaker, is linguistic creativity.

Based on the boundaries established above, the relationship between linguistic creativity and linguistic innovation can also be analysed as being located on different levels of language use. Linguistic creativity happens on the level of the idiolect, whereas linguistic innovation has to transcend that level and be located on the communal language level. As Mufwene (2008, 2) puts it:

> No speaker plans to change their idiolect or communal language through the innovations and deviations they produce during their communicative acts. Since they do not all do the same things at the same time, nor do they consult with each other about what to say (or how to express some new ideas), it is interesting to at least take note of the intriguing phenomena of how patterns emerge and spread into norms from these innovations and deviations.

The various aspects and consequences of linguistic agency are part of a discussion occurring in the field of Evolutionary Linguistics. Before we take a closer look at agency itself, and the role it plays in linguistic creativity, it will be necessary to establish the central tenets of this field of linguistics.

2.3 Evolutionary Linguistics, complexity and Usage Based Grammar

2.3.1 Croft's foundations

Many of the concepts and processes discussed in Evolutionary Linguistics go back to Croft (2000), which may be seen as the foundational work in the field. One of his fundamental assumptions is that "[w]hen linguists analyse language as an abstract system, they are not looking at a historical entity, nor are they looking at a type about which predictions can be made" (Croft 2000, 2). Rather, he claims, the object of a linguist's study then is "not real at all, either as a type or as a token", which casts doubt on the status of linguistics

as an empirical science, which starts with observations and builds generalizations and models based on these observations. To remedy this conundrum, one needs to focus on utterances as the actual empirical objects of study on the one hand and on the speaker who, using their minds and knowledge of language, produce these utterances on the other. Only by working outward from these utterances can more complex constructions and patterns, such as languages as a whole or a speech community of more than one speaker be approached.

Croft's work is largely concerned with language change. He identifies two fundamental processes in linguistic, or rather, any kind of change: inherent change and replication. The first, he defines as "a single object that exists over time chang[ing] in some way or other" (Croft 2000, 3), using the physiological development of humans as a biological example, but likening it to human linguistic development, or "the development of mental structures that we interpret as representing linguistics knowledge, over the lifetime of the speaker". Replication, on the other hand, is defined as "the creation of a new entity that preserves in large part the structure of its parent entity (or entities)". Biologically, this would include DNA and meiosis, whereas linguistically, it involves any utterance, which, by necessity is based on previously existing grammatical structures, which have in turn been established by previous utterances of the same and other speakers.

Croft goes on to postulate five desiderata for a theory of language change, the first of which states that "a theory of language change must avoid the reification or hypostatization of languages" (Croft 2000, 4). It is not language that engages in the processes which ultimately lead to changes within them, but speakers. In Croft's words, "languages don't change, people change language through their actions" (Croft 2006, 4). The agency - a concept that we will return to in some detail below - lies within individuals, and it is in their actions and motivations that explanations for language change can be found. "Secondly", Croft continues, "a theory of language change must explain why languages do NOT change in many ways, sometimes over many generations of speakers" (Croft 2006, 4). He rightly points out that if factors which cause language change were the only ones asserting their influence, languages would be in a lot more flux and subject to constant change. Since they are not, "a theory of language change must provide mechanisms of NORMAL (identical) REPLICATION as well as altered replication". Croft's third desideratum is that any theory of language change should distinguish between altered replication and differential replication. In linguistics, he asserts, these are "INNOVATION or actuation - the creation of novel forms in the language - and PROPAGATION or diffusion (or, conversely, loss) of those forms in the language" (Croft 2000,

4-5). He recognizes both as subprocesses of a larger overall process of language change, noting that the distinction between both is "rarely recognized in models of language change" (Croft 2000, 5). The fourth desideratum Croft puts forth is that "a comprehensive framework for understanding language change must subsume structural, functional and social dimensions of language change, or their equivalents", criticizing that many contemporary theoretical approaches are too focused on one of these aspects, while leaving out the others. Finally, Croft's desired theory also "must subsume both INTERNAL and EXTERNAL causes of language change".

Going into more detail on the processes of innovation and propagation, Croft explores the causal factors underlying them and how they differ:

> The causal mechanisms for innovation involve the mapping from language structure or form to language function, that is, meaning in context [...]. This mapping occurs at two levels or interfaces. One is the mapping from phonological structure to phonetic reality (articulatory and auditory). The other is the mapping from grammatical (morphosyntactic) structure to its semantic/pragmatic/discourse function in context. Neither mapping has been found by linguists to be simple to represent. But nor is it simple for speakers to represent these two levels of mapping. [...] I argue that altered replication is essentially a result of speakers adjusting the mapping from language structure to external function. [...] The mechanisms for innovation in language change involve both structure and function. The mechanism for propagation, on the other hand, are essentially social, namely the various factors discussed by sociolinguists [...]. In other words, there are two distinct mechanisms operating in language change [...]. The mechanism for innovation is functional, that is, involves the form-function mapping. The mechanism for propagation is a selection mechanism, in the evolutionary sense [...], and it is social. Thus, the integration of structural, functional and social dimensions of language change is achieved largely by integrating the two distinct processes of change, innovation and propagation/selection. (Croft 2000, 8)

Innovation, Croft claims, does not occur at arbitrary spots within the system of language. Rather, it happens at "cracks in the system of conventions to which speakers are attempting to conform when speaking" (Croft 2000, 117). One of these cracks, he argues, is the impossibility to separate between the established use of a grammatical form within a speech community in terms of context and relations to other forms from a nonconventional use of the form in a specific context. Speaker and hearer both, in turn, rely on their existing grammatical knowledge and previous exposure to the form in order to convey and deconstruct, respectively, the intended meaning in a particular context.

To this, they add knowledge about the current context in order to ensure successful communication. This entire process would be unnecessary if not for the ability of language users to recombine words and constructions, which, according to Croft, "clearly adds a further layer of complexity onto the process of using language" (2000, 117). The benefit, of course, is that by recombining existing elements, a theoretically infinite set of meanings can be constructed and applied to the current situation.

This raises the question of how speakers acquire their knowledge in the first place. According to Croft, "regularities [...] must be abstracted from histories of language uses in their rich communicative context" (Croft 2000, 118). Such abstraction, along with analysis, are the fundamental processes underpinning language use in his theory. He notes that this applies to both hearers/learners as well as speakers, all of which rely on abstraction and analysis of utterances they have previously been exposed to in order to construct new utterances whose meaning can be successfully decoded by other interlocutors. In learning a new language, we are not exposed to analytical parts of it, but to complete utterances that must be divided (analyzed) into their components - "syntactic and semantic", as Croft (2000, 118) notes - in order for abstraction to work. By abstracting the meaning of particular constructions, "a mapping from syntactic units onto components of meaning in the speaker's mind" is constructed. However, as there are likely to be subtle differences in each communicative context, "the recombination in utterances of pre-existing grammatical units and structures will involve some degree of novelty in the form-to-function mapping in each use of language", somewhat akin to the concept of 'little c' creativity discussed above. Since both sides of the form-function equation are complex units, and "since the construction is a fixed whole, and so is the semantic structure it denotes, the syntactic units and their semantic denotata repeatedly co-occur" (Croft 2000, 118). The resulting ambiguity from the same form-function pairings occurring in different contexts provides "leeway", as Croft terms it, for form-function-reanalysis to occur:

> Form-function reanalysis is a nonintentional mechanism for innovation. Speakers' intended actions are towards conformity to convention, but the result is innovation, an unintended consequence. The unintended consequence of individual actions is due in part to the potential discrepancy between individual competence, which is constantly responding to use [...], and the conventions of the speech community. Most of the time, the grammatical forms produced are essentially the same as those produced before, albeit in novel combinations, in novel meanings in context, and also with variable pronunciations [...].
> (Croft 2000, 118f.)

Croft underpins this theory with empirical data from Bybee and Slobin (1982, 268), who observed their participants produce new regularizations and novel stem forms under the pressure of time when asked to provide the past tense forms of 90 irregular and three times as many regular verbs. Since the ninety participants were explicitly trying to produce correct forms - not innovative forms - Croft takes this as evidence that even when following norms and relying on previous knowledge, speakers produce aberrant forms - which can be seen either as errors or as innovations. He argues that the innovations observed by Bybee and Slobin are "likely to be a stochastic outcome of spreading activation and interactive activation in the network of linguistic knowledge, in this case, the knowledge of morphological forms and their meanings" (Croft 2000, 119). A similar process could be responsible for form-function reanalysis when grammatical constructions and semantic constructions are exposed to interactive activation.

The actual mechanism underlying such form-function reanalysis, Croft (2000, 119) asserts, "is based on the form-meaning mapping in grammatical constructions" and thus "on a very simple model of grammatical structure", in which "an utterance represents a construction that is made up of its component syntactic, lexical and morphological units; the grammatical units have semantic values [...]; and the semantic values of grammatical units contribute somehow to the semantic value of the syntactic constructions of which they are a part". And that is when the abduction happens:

> Form-function reanalysis is abductive [...] [,] that is, the representation of the form-meaning mapping in the speaker's and listener's heads is abducted (inferred or reconstructed) from prior and current experiences of the form-meaning mapping found with similar utterances in similar situations. This abduction occurs in language use, however, not in language acquisition [...] thereby conforming with what we know about how language change actually occurs.
> (Croft 2000, 119)

Merely innovating new form-function pairings, however, is insufficient for lasting language change. To achieve such a goal, the form-function pairing has to be spread across the speech community as a whole on one hand and possibly come into competition with other form-function pairings competing for the same function. This is where additional aspects of Evolutionary Linguistics come in: those of competition, selection, and the feature pool as proposed by Salikoko Mufwene.

2.3.2 Mufwene's theories

Competition and selection, feature pool

Mufwene's main - or at least, best known - contribution to the theories of Evolutionary Linguistics is that of the feature pool. This notion, as he says, is based on two main working assumptions, namely that a) "the agency of the evolution of a language lies in the individual communicative acts of its speakers, just like a biological population or species is cumulatively affected by the experiences or activities of the individuals it consists of" (Mufwene 2008, 11) and that b) "the communicative activities that produce language evolution are largely determined by the socioeconomic ecologies in which speakers evolve, which is similar to saying that the ecology rolls the dice in evolution". From these two assumptions follow several important conclusions. First, if the agency of linguistic evolution lies with the individual speaker, that must mean that any speaker is capable of linguistic creativity - be it consciously or unconsciously. It is by the summary of the linguistic actions of the speakers that a speech community behaves, and it is therefore by the summary of the linguistic changes individual speakers make to their speech behaviour that changes happen in the speech community. However, from b) it follows that speakers make no random choices and are not free in their speech behaviour. Instead, it is their circumstances and their communicative purpose which shape their individual behaviour, which in turn shapes the group's behaviour.

The feature pool, then, is the "total amount of linguistic knowledge" (Mufwene 2008, 17) that an individual speaker may draw from in order to shape her own linguistic output:

> From the point of view of the development of linguistic or communicative competence, the total amount of linguistic knowledge speakers have, including the variants that compete for the same structural or communicative functions, is comparable to a "gene pool" in biology, both at the individual and at the population levels. In the case of language, the term feature pool [...] seems to be an apt analogue. (Mufwene 2008, 17)

In her feature pool, a speaker may find different forms that fulfil the same function, i.e. variation is an inherent part of the concept. As Mufwene puts it in the quote above, these different variants compete for the same function. The outcome of this competition is not decided randomly: some variants may receive an advantage due to analogous other features, some may be more entrenched, others may be more morphologically transparent. In connections

with contact languages, some may work similarly or the same as in a source language, which would be another benefit. If one were inclined to stretch the pool metaphor to its limit, one could say that certain features swim more closely to the surface of the pool and are more easily accessed (and spread), whereas others are dragged further down by characteristics that put them at a disadvantage.

The fact that an individual speaker and her idiolect are capable of holding different variants - be they lexical variants, phonological variants or grammatical variants - in their feature pool also has a mirror in evolutionary biology, namely polyploidy. In biology, this is the state of a cell having more than two paired sets of chromosomes. In theory, linguistic polyploidy is unlimited, but the speaker will still select some and discard others. This selection happens entirely without agency of the gene inheritor in biology, whereas in linguistics, the speaker can make a conscious, agentive choice for a specific variant. The theoretical agency of the speaker is higher than of the biological inheritor, but this does not mean that it is always necessary (Mufwene 2008, 19). To quote:

> [I]t is not clear what factors or what particular selection algorithm a biologist would invoke to account for why particular genes become dominant or recessive in the genotype of a particular organism or for why a particular individual winds up with a particular combination of phenotypes. Linguists can invoke factors such as the statistical frequency of a feature, semantic transparency, regularity, salience, and social status of the model speakers. (Mufwene 2008, 19)

Any of these factors can make one variant more competitive and more likely to be selected when agency is not the primary factor in selection. A speaker may choose to deliberately use a factor that is less frequent, semantically less transparent, less regular, less salient and connotated with less social status, but usually only when consciously doing so. Otherwise, these factors will serve to order the variants within the feature pool and make one more competitive than the others.

This selection, Mufwene claims, happens on the levels of features (i.e. units or their combination), not at the level of languages. Languages are only indirectly selected through their features. This is also an important process in the creation of contact languages. If speakers opt - again, consciously or unconsciously - to adopt a majority of the features of one language, then the emerging contact language will "wind up constituting the majority of those selected from the combined feature pool of the language varieties in contact" (Mufwene 2008, 20). Since in some contact settings - those, according to

Mufwene, that produce creoles - the evolutionary advantages of the features of one language are so high, "only one target language, often misnamed the super-stratum or lexifier, lends most of its vocabulary and grammar to the emergent vernacular" (Mufwene 2008, 20). However, he sees the fact that features of the substrate languages are selected as well as proof that the selection happens on a feature level rather than a language level. If the latter were true, then no substratum features should be selected at all.

With regard to which features are selected, Mufwene notes that both internal and external ecologies of linguistic interaction apply. This is in concordance with many models of language change in which both external factors and internal changes drive the process. Mufwene additionally establishes a similarity to epidemiology, "as social practices bear on how viruses spread in a population, although different species are subject to different constraints specific to them in the selection of their features" (Mufwene 2008, 20).

In terms of language change, however, which changes take place and how they spread are only two of a three-set part of questions. The third is the question that Weinreich et al. (1968, 102) have termed the actuation problem. If the other two can be reduced to "what changes" and "how does it spread", this question is the one of "why now/then"? What factors have pressed the speakers into adding, removing or changing elements of the language system at this particular time? And why does the same change happen in different languages at different stages? Note that what the actual actuation question is differs among linguists. In McMahon's (2002, 248) view, for instance, "the real actuation question is why some of these innovations [...] die out and others catch on, spreading through the community, or why certain instances of variation become changes while other's don't". In my understanding, actuation and diffusion are two different subprocesses of language change: the former I understand to be the actual first instance of a change in one speaker's (or, in the case of parallel innovation, multiple speakers') idiolect(s), whereas the latter is the spreading of said change across the speech community as a whole.

Actuation, then, is a process that starts with the individual speaker and her idiolect. This brings back the question of how conscious the individual speaker is of changes she makes to her own speech, which may latter diffuse across the speech community, which, finally, brings us back to the concept of (linguistic) agency as mentioned in the overview of linguistic creativity above.

The invisible hand and the agency of the individual speaker

Duranti (2007, 453) understands agency to be the "property of those entities (i) that have some degree of control over their own behavior, (ii) whose actions in the world affect other entities' (and sometimes their own), and (iii) whose actions are the object of evaluation (e.g. in terms of their responsibility for a given outcome)". In linguistic terms, we might say the property of a speaker to control her own speech, to affect others' speech and to have her speech evaluated by other speakers. Duranti's first point also infers that to have agency, a speaker must be conscious of the changes she makes to her own speech.

As a provisional definition, Ahearn puts forth the notion that "agency refers to the socioculturally mediated capacity to act" (Ahearn 2001, 112f.), though she immediately acknowledges that while this definition serves as a starting point, "it leaves many details unspecified". These unspecified details include, according to her, the answers to the following questions:

> Must all agency be human? Can nonhuman primates (Small 1993), machines (Pickering 1995), technologies (Dobres 2000), spirits (Keane 1997, pp. 64–66), or signs (Colapietro 1989, pp. 95–97; Peirce 1955) exercise agency? Must agency be individual, leading to charges of unwarranted assumptions regarding Western atomic individualism (Ortner 1996)? Or can agency also be supraindividual—the property, perhaps, of families, faculties, or labor unions? Conversely, can agency be subindividual—the property of "dividuals" (Daniel 1984, p. 42; Marriott 1976; McElhinny 1998, p. 181), as when someone feels torn within herself or himself? What does it mean to be an agent of someone else? Must agency be conscious, intentional, or effective? What does it mean for an act to be conscious, intentional, or effective?

The linguistic transfer of these questions would then be whether agency can only be ascribed to the individual speaker, or to the speech community as a whole on the one hand and whether speakers who make unconscious changes to a language lack agency. Among other approaches, this question is handled in Evolutionary Linguistics through the 'invisible hand' phenomenon.

The so-called 'invisible hand' phenomenon has its origins not in evolutionary biology, as many of the other notions Mufwene adopts, but in economics. The expression was coined by Adam Smith in 1776 in his fundamental work *The Wealth of Nations*. Smith described it as the unintentional side effect of an individual's decision on a larger scale. The individual "intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention" (Smith 2007, 349). Mufwene

(2008, 60) sums up the concept in its relation to linguistics and language change as follows:

> In the case of language, the 'invisible hand' amounts to the cumulative, though typically uncoordinated, actions of individual speakers which bring about change, this being typically an unintended outcome. The focus in this case should be both on the individual and on the role of imitation in group behavior. As some speakers adopt innovations or deviations produced by other speakers, new structural and/or pragmatic patterns emerge. As more and more new linguistic behaviors similar to the initial innovation occur, new norms emerge at the communal level and are identified as changes relative to an earlier stage of the relevant language. Norms emerge from the convergence of speakers' linguistic behaviors or from some speakers copying innovations or deviations produced by others in more or less the same way that footpaths emerge from the footsteps of individual pedestrians following more or less the same trajectory.

There are various aspects to this definition that warrant further explanation, such as the dual nature of the invisible hand. First of all, the 'invisible' aspect of the 'invisible hand' refers to the fact that language change and development happens mostly unconsciously. As such, it is closely related to the question of agency, which we will deal with in detail below. The central point here is that the 'hand', whatever it may be, does not steer the speakers actively. Instead, the picture of the hand only emerges in retrospect, when a certain number of speakers seems to have chosen the same path. Note that this does not have to include all the speakers all of the time, or even some speakers all of the time. There is no authority or concert required over or between speakers for them to make identical or similar decisions. In other words, the speech community as a whole requires no agency, it is the speakers employing similar coping strategies when faced with the same communicative challenges that lets patterns emerge. According to Mufwene (2008, 61), "phenomena known as grammaticalization evolve this way, with innovations by some speakers spreading through copying within a population". In this view, Mufwene aligns with other theories of language change such as the 'invisible hand' conception of Rudi Keller, who sees language change as neither a man-made phenomenon nor a natural phenomenon but a causal consequence of the actions of many individual actions (Keller 1994, 51-57).

It is worth noting that this ties back to the notions of selection and competition in evolutionary linguistics: an individual speaker comes up with an innovation, that may get selected by other speakers as a suitable tool to solve a communicational problem and then spread through the rest of the population,

likely coming into competition with previously existing solutions innovated by other speakers at earlier times or at the same time. Naturally, it is also possible for multiple speakers to come up with the exact same solution at roughly the same time, in a case of parallel innovation, which raises the question of whether such solutions are somehow more 'natural' or more suitable. Coming up with different solutions, however, is part of the reason that languages can be variable. Other reasons include the following, as put forth by Mufwene (2008, 62):

1. a communal language is an extrapolation from the idiolects of its speakers - it can be conceived of only on the basis of experience with some idiolects;

2. a language is internally variable, in part because its idiolects are not identical with one another - neither the experiences on the basis of which they have evolved nor the minds that have developed them are identical; and

3. a language does not evolve in a uniform way, because each speaker's communicative network is identical with nobody else's and each speaker's communicative experience is unique.

He also notes that evolutionary patterns emerge along various social groups, be they defined by age, gender, education, professional occupation, etc. However, he only regards these as "trends, not mathematical categories" (Mufwene 2008, 62). Since not all idiolects are affected by or involved in all changes a language undergoes, and even those involved may do so at different speeds, it is only in the more abstract developments that we can locate trends and speak of the development of a language or speaker community as a whole.

This brings us to the next point: how to see an individual speaker and her idiolect in relation to the speech community and its language as a whole, in evolutionary terms. Mufwene puts forth the idea that specific languages should be seen as species, while idiolects - not speakers - should be seen as the species' members (Mufwene 2008, 63). It is the idiolects, after all, which resemble each other on a structural level, akin to the phenotypes of biology, not the speakers themselves. In addition, he argues that "speakers of similar idiolects also acknowledge that they speak the same language variety and they often assume that it descends from an earlier form to which their idiolects can be genetically related".

Mufwene relates to Wittgenstein's notion of family resemblance here. The basis for modern prototype theory, it theorizes that certain entities need not share one essential feature to be connected, but instead may be connected by a

series of overlapping features. It is on these superficial, or structural features, that idiolects exhibit enough matching material to be grouped into one species, or language.

When Mufwene talks about languages as species, he understands them to be Lamarckian species. The main difference between the Lamarckian and Darwinian concept of evolution is that the former centres on change through use and disuse, rather than (more or less) random genetic mutation. Lamarckian evolution assumes that that as environments changed, organisms changed their behaviour according to the new environment and their wants and needs, the latter factors not being relevant in Darwinian evolution Burkhardt (2013). While Darwin's theories worked better in evolutionary biology, Lamarck's concept serves better in explaining cultural evolution, including linguistic change. Mufwene (2008, 65f.) claims that linguistic species are Lamarckian because they are subject to change on more than one occasion during their lifetimes. Unlike the genetic structure of a singular biological organism, which remains stable throughout its lifetime, variation is possible even within the same idiolect. This also has implications for the notion of family resemblance among idiolects, because it is not inherited family resemblance, but rather acquired family resemblance. In other words, family resemblances between two idiolects may be lost as one speaker radically changes her idiolect.

How then, do these individual members of a species interact and shape the species as a whole? This is where Mufwene comes up with a somewhat unusual simile that unexpectedly works quite well: that of population-wide patterns in language with highway traffic.

> I wish to highlight similarities between the way individual speakers influence the evolution of their communal language and how the actions of individual drivers on a highway influence traffic flow. This comparison is in keeping with my assumption that the way in which a communal language changes is a function of how individual idiolects change under each other's influence and as a cumulation of adjustments that take place during their individual speech acts. Likewise, the way traffic flows on a highway is largely a function of how individual motor vehicles proceed. It is also true that what individual vehicles can do depends on the overall traffic itself, such as its volume. Likewise, the changes that idiolects can undergo are largely constrained by the language they are part of. That is, what individual speakers can produce and how they can modify their idiolects to express new ideas depends largely on what other speakers of the same language variety do or can understand. (Mufwene 2008, 59)

If we entertain this notion (and accept that the metaphor somewhat conflates

the speaker and her idiolect, unless we imagine the metaphor to describe them as the driver and the car, respectively), we might find that the analogy works better than expected: the overall traffic directs some of the actions of the individual driver or speaker. She can mostly affect those drivers or speakers directly around her, i.e. those in immediate contact. However, changes at one point in traffic can have a ripple effect and cause a traffic jam - or smooth traffic - miles away. The point at which she enters the highway traffic, and which drivers or speakers she comes into contact with can influence a speaker's driving/idiolect. Speakers can also leave the highway, exiting a particular speech community in favour of others.

However, Mufwene also mentions a very important point in language change here that I want to highlight: even though it happens on the level of the member/idiolect first, it still is constrained by the species/language as a whole, since there is a need for transparency. While it is theoretically possible to change our idiolect in any way we wish to - as long as we are conscious of what we would have to change - the change is very unlikely to spread if it is not transparent to other speakers, even if their idiolect belongs to the same species/language. This ties back in with the question of agency. We have to keep in mind that despite the fact that speakers work under constraints when changing their own idiolects, it is still the speakers themselves who drive the change, not the idiolects. As Mufwene (2008, 64) puts it, "[j]ust as vehicles on the highway cannot move without the agency of drivers, languages do not evolve without the agency of speakers, on whom their lives depend".

The notion is this: we do not drive around randomly in traffic, and neither do our fellow drivers, despite what a Monday morning traffic jam may make us believe. Traffic is not random, especially not when seen on a larger scale. Neither is language evolution, since speakers, too, usually do not go around changing their idiolect on a whim. If we accept this premise, then it follows that language change and language evolution do not happen on a path devised by internal language structure, but are shaped by the language-external needs and restraints of its speakers. While this may be obvious to today's linguists, especially sociolinguists, it is worth remembering that this conviction has not always been held. Going back to Schleicher's (1863, 1869) work, there used to be a notion, especially prevalent in the nineteenth century, of an evolutionary trajectory of language, akin to the evolutionary trajectory ending in the human species in biology. The unfortunate Eurocentrism in this notion led to languages with the fusional morphology type being considered more complex and more evolved when compared to languages of the isolating morphosyntactic type, despite the fact that the latter includes languages

such as Chinese. As Mufwene (2008, 12) laments, this misguided notion also led to early research into pidgins and creoles being carried out under the illusion that they were primitive languages, created under circumstances which reversed the evolutionary processes that had led to the higher stages of the colonial languages. Fortunately, that view is no longer prevalent, and "[e]xcept for Bickerton (1984a, 1990), who, through his language bioprogram hypothesis, has suggested that creoles and pidgins give us an idea of the human protolanguage, no language typologist has ever suggested any particular evolutionary ranking of structural types" (Mufwene 2008, 12).

If we view languages as Lamarckian species, whose evolution is driven by individual organisms, that is, by speakers' idiolects, then it necessarily follows that changes in the environment of these idiolects - new contexts and new information they need to convey, for instance - affect linguistic development. In general, such environments can become more or less complex over time. They are then expressed by qualitatively and quantitatively more or less complex ways and it is this process which is the causal factor in the species/language itself becoming more or less complex. But what does it mean for a language to become more or less complex? And how does this complexity eventually express itself in its surface structures?

2.3.3 Complexity, adaptivity and usage-based grammar

The first step in answering these questions is to define what we understand when referring to complexity in linguistics. The answer may appear obvious: as Mufwene et al. (2017, 1) note, "one is [...] shocked by the scarcity of works that explain what COMPLEXITY is, apparently because it is assumed to be known". Drawing on other scientific fields, they come up with the following features and themes that are relevant to complexity (ibidem):

1. Complexity arises from the coexistence of components that interact with each other, not necessarily from the fact that a space or a system is populated with several components or members; it is therefore interactional.
2. Complexity arises from the dynamics of activity coordination or synchronization that integrate individuals as members of a population (e.g., ant colonies, bird flocks, and fish schools); thus, it is dynamical.
3. Complexity emerges from nonlinear evolution, which is driven by multiple factors whose significance may vary at different stages of the evolutionary process; its effects are not constant, but subject to the changing values of the relevant variables.

4. Complexity lies in what brings order out of chaos, through what is also known as 'self-organization' and was formerly referred to as an 'invisible hand'.

5. There is complexity in any system where the properties of the whole do not amount to the sum of the properties of the components.

6. Finally, complexity is the peculiarity of emergent patterns in a system in constant state of flux between disorder and transient order (or equilibrium). In other words, complexity arises from the dynamics of coexistence and interaction or cooperation of components toward generating the properties of whole.

All of these features are relevant in complex linguistic systems as well, with the first one being the most obvious: a system is complex if it consists of several interacting components. This is the case for multiple levels in linguistics. On a higher level, we have phonetics/phonology, lexis and morphosyntax interacting with each other. They, in turn, consist of several components which interact as well. One additional important notion that Mufwene et al. mention is that "complexity arises not just from how the different parts interact with each other but also from how they respond to external pressures of the environment" (2017, 1). These external influences do not only affect the more abstract levels like morphosyntax, but also their individual components, as do the internal forces. So we end up with a linguistic system that is inherently complex, with parts dependent on and influencing each other on various levels, and then add external pressure on every level, which results in even more interactions.

A second layer of complexity stems from the dynamicity of the system, i.e. the interactions within the system can lead to change within the system. Again, if we continue to pursue the notion of languages being species and the individual members being idiolects, this would indicate that needs and wants of speakers are external factors, and the system is dynamical because the speakers change their idiolects, which in turn change the system they are part of. So rather than an individual ant in an ant colony, individual members of the species reflected in a linguistic system have no agency. This also illustrates the problem of whether to conceptualize each idiolect as its own system, with the system of the entire language or the entire speech community as a reflection of the more abstract patterns within the individual idiolects, or to see idiolects as part of the same system as the language itself. I would argue that the latter is more beneficial, given that contact between idiolects plays a role in any change in either system.

The third point - complexity emerging from nonlinear evolution in which factors can have different impacts throughout time - is rather simple to apply

to language as well. For one, external factors surely are subject to change throughout time: contact with other languages and other speech communities, for instance, can differ in how intimate it is, what the differences in prestige are, etc. Changes in the environment or culture can cause parts of the system to become more or less relevant due to changing needs and wants of the speakers the idiolects are tied to. Language-internal factors may change as well. Since language change is inevitable, be it caused by external factors or not, parts of the system will change. As parts of the system change, this will affect the parts they interact with, possibly causing them to change as well.

We have already discussed the notion of the 'invisible hand' above, in chapter 2.3.2. Certainly, there needs to be a certain amount of order within a linguistic system - i.e., somewhat stable components and somewhat stable interactions - to ensure the necessary transparency in order for it to be used for communication. In order to explain whether or not the 'sum' of a linguistic system is greater than its parts, we will first have to determine what we consider its sum to be. Mufwene et al. (2017, 4) first seem to consider this on the level of individual items or constructions:

> A good example is when an item generates different interpretations depending on what other item it is combined with. This is illustrated with the particle up in combinations with various verbs such as in pick up, give up, show up, and look up. While the item *up* is basically the same particle in all these constructions, its contribution to the meaning of each phrase appears to vary. This variation suggests that the particular dynamics of each combination produce the meaning of the whole phrase.

In other words, the meaning that derives from the structure of the parts is something that goes beyond the sum of the parts themselves, since it also includes their relation to each other. The notion is also equally valid on several other levels of a linguistic system and could even be said to go back to the Saussurean notion that signs derive their value from their position within the system and relation to other parts of the same system. Alternatively, if we see phonology, morphosyntax and lexis as the structural parts of the system, then we could consider semantics and pragmatics as the 'additional' part that emerges when adding up the structural features.

The last feature Mufwene et al. posit boils down to the fact that the system becomes more complex as order within the system is not permanent since its parts are always changing both in their inherent structure and their interaction with each other. It is this dynamic interaction that generates the systems' properties, not a stable set of rules. They mention the state of 'equilibrium',

which raises the question of whether a decrease in complexity within one part of the system leads to an increase in another in order to keep a putative overall complexity level the same. Consider the following:

> Such a conception of equilibrium can also be considered within a domain, if, for example, one attempts to check whether the greater complexity of the consonant or vowel system is counterbalanced by the lesser complexity of the syllabic structures (Maddieson 2011). Technically, comparing distinct domains such as phonology and morphosyntax is uneasy beyond simply counting elements, which, as remarked earlier, typically disregards the interactions between them. (Mufwene et al. 2017, 6)

It is not an uncommon notion in linguistics that changes that affect one subsystem of language can affect another as well, such as higher case syncretism leading to a more rigid word order to ensure identifiability of syntactic constituents. Neither is it a new notion, going back at least as far as Sapir (1921, 66) and Jakobson (1936, 28).[2] As Blake (2001, 15) notes, "it has frequently been observed that there is a correlation between the presence of case marking on noun phrases for the subject-object distinction and flexible word order and this would appear to hold true". Again, we are facing a multi-level system that is interconnected on all levels: complexity of individual elements can be 'balanced' between components of the same subsystem, e.g. phonology, as in Mufwene's example above, or it may be balanced across subsystems, as is the case in the case and word order example. Conversely, this means that a change anywhere within a subsystem can have wide-ranging effects on the same subsystem or other subsystems that it is interacting with.

We have now established the factors that make language a complex system of many interactive parts, but there is at least one additional aspect to consider when describing language in systemic terms, namely that of it being adaptive. We have already mentioned that part of the complexity of the system as a whole arises from the fact that changes within one part affect changes as a whole. The very possibility that changes may occur within the linguistic system is due to it being an adaptive system, i.e. one that responds to internal and external pressure by changing its structure.[3]

---

[2] The former notes that in languages with high case syncretism, we "we cannot afford to be so indifferent to our word order. We need to husband our resources. In other words, word order takes on a real functional value", while Jakobson observes that in inflectionless languages, it is the word order which determines the function of the nouns ("[...] dort aber, wo die Kasusform unklar ist [...], wird die Funktion der Nomina im Satz durch die Wortfolge bestimmt".

[3] It is possible for complex systems to be non-adaptive, i.e. rigid. This would describe a system whose individual subsystems and components interact with each other, but

The conception of language as a complex and adaptive system was pioneered by the Five Graces Group in their position paper 'Language is a Complex Adaptive System'.[4] Their core position is as follows:

> Language has a fundamentally social function. Processes of human interaction along with domain-general cognitive processes shape the structure and knowledge of language. Recent research across a variety of disciplines in the cognitive sciences has demonstrated that patterns of use strongly affect how language is acquired, is structured, is organized in cognition, and changes over time. However, there is mounting evidence that processes of language acquisition, use, and change are not independent of one another but are facets of the same system. We argue that this system is best construed as a complex adaptive system (CAS). This system is radically different from the static system of grammatical principles characteristic of the widely held generativist approach. Instead, language as a CAS of dynamic usage and its experience involves the following key features: (a) The system consists of multiple agents (the speakers in the speech community) interacting with one another. (b) The system is adaptive; that is, speakers' behavior is based on their past interactions, and current and past interactions together feed forward into future behavior. (c) A speaker's behavior is the consequence of competing factors ranging from perceptual mechanics to social motivations. (d) The structures of language emerge from interrelated patterns of experience, social interaction, and cognitive processes. (Beckner et al. 2009, 2)

We have already established that it is primarily point (a), the fact that the system consists of a multitude of interacting agents (speakers in the CAS, idiolects in Mufwene's model), which makes the system complex. The competition point (c) refers to has been covered as well (see chapter 2.3.2). Point (b), however, should be of equal interest. It is not just that the system of a language is subject to change from external forces, such as new requirements for expression and communication, but the idiolect is influenced not only by current circumstances. Instead, it is shaped by past interactions as well. Since the idiolect is the individual component of the system at large, and the idiolect is in turn shaped by the past interactions between idiolects - factors such as the frequency at which the idiolect's speaker was exposed to certain constructions, for instance - any language use in the past would shape language use in the future. If we assume that frequency of exposure plays a role in governing the structure of idiolects and ultimately, a language as a whole, then this, at

---

neither their internal structure nor their interactions are subject to change. As such, this kind of system is a more useful notion in engineering and mechanics than in cultural evolution, where systems are assumed to be responsive to change and therefore adaptive.

[4]Note that the Five Graces Group, despite its name, contained more than just five scholars.

the very latest, is the point at which it becomes obvious that the notion of language as a complex adaptive system requires a usage-based approach, "in which the cognitive organization of language is based directly on experience with language" (Beckner et al. 2009, 5). An abstract set of rules or structures that are inherently governed and not directly influenced by its use would not be compatible with an adaptive system.

The core tenets of usage-based grammar are summarized by Zeschel (2012, 11) as follows:

> ...usage-based linguistic theories assume that speakers' grasp of a language arises from their categorisations of concrete linguistic usage events, where the term 'usage event' is defined as follows: "the pairing of a vocalization, in all its specificity, with a conceptualization representing its full contextual understanding. A usage event is thus an utterance characterized in all the phonetic and conceptual detail a language user is capable of apprehending" (Langacker 1999: 99). Speakers are assumed to analyse and sort the structure of these experiences by mapping particular aspects of the input to relevantly similar elements of long-term memory (in several dimensions in parallel, thus recognising particular morphemes, words, multi-word units, syntactic phrases, argument structure constructions, intonation contours etc.). For this, linguistic categorisation draws on a powerful capacity for pattern matching in human cognition that is not peculiar to language. Likewise, speakers' internalised linguistic systems are assumed to 'emerge' from countless individual categorising events of the abovementioned type in a way that does not presuppose any domain-specific innate constraints on possible grammatical abstractions. Second, the resulting system is nevertheless assumed to be highly structured: metaphorically speaking, new elements are stored 'next to' similar pre-existent units and contract all sorts of connections to other stored elements on the basis of perceived similarities (Bybee 2006). With growing exposure, speakers thus develop an increasingly complex, network-like structured inventory of categorised symbolic assemblies that exhibit varying degrees of entrenchment (cognitive routinisation). And third, the overall inventory is seen as fluid and dynamic since it is constantly adapting to experience. Put differently, usage-based theories assume that speakers' internalised linguistic system does at no point settle to a more or less unchanging 'final state' as assumed in Chomskyan theories of language acquisition, and that this system will also vary from one speaker to the next in many respects (notably when it comes to less salient properties of less frequent constructions whose representations are not constantly aligned and hence accommodated accordingly).

Having established usage-based grammar as the basis, we come back to point (d) of language as a complex adaptive system, the claim that "the structures of language emerge from interrelated patterns of experience, social interaction and cognitive processes". One model based on such a notion is Schmid's EC (Entrenchment and Conventionalization) model, which is detailed in figure 2.2. Schmid (2015, 3) sums its components up as follows:
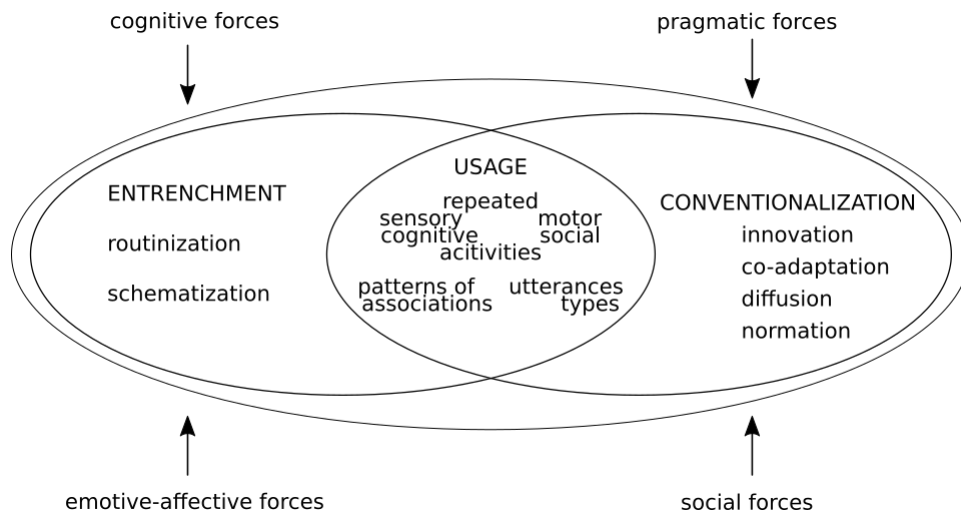
Figure 2.2: Entrenchment and Conventionalization Model

- linguistic usage and four types of repeated activities[5] involved in it

- a limited set of cognitive processes operating in the minds of speakers, subsumed under the label *entrenchment*: *association*, *routinization* and *schematization*

- a limited set of sociopragmatic processes operating in communities, subsumed under the label *conventionalization*: *innovation*, co-*adaptation*, *diffusion* and normation

- a (probably unlimited) set of cognitive, emotive, pragmatic, and social forces which influence the way in which *entrenchment* and *conventionalization* processes interact with usage to shape and change language

As can be seen in Figure 2.2, schematically adapted from Schmid (2015), the three individual parts of the model are not separate, but overlap to some

---

[5]These four types are the activities Schmid sees as central to language use, namely motor activity (the production of utterances, regardless of mode; sensory activity (the perceival of utterances and the context); cognitive and neuronal activity (planning, formulating and understanding utterances in context); and social and interpersonal activity bound to be part of communication.

extent, indicating that "usage affects both entrenchment and conventional-ization, while entrenchment and conventionalization in turn influence usage" (Schmid 2015, 7). The outer ellipses, however, have no direct link and are instead only connected through usage, with no factor or process of entrench-ment directly affecting conventionalization or vice versa since "only usage in interaction affords the constant updating of individual cognitive and collective social systems". The external forces outside the outer ellipsis are arrayed by which process they have the most direct influence on - cognitive and emotive-affective forces on the side of entrenchment, and pragmatic and social forces on the side of conventionalization. Note, however, that this only means they affect this process more directly and strongly, while "all four types of forces can influence the entire interaction of the three core components" (2015, 7).

For the present endeavour, however, it is the three types of cognitive pro-cesses which underlie conventionalization that are of most interest. Conven-tionalization can be linked to the diffusion concept in evolutionary linguistics (see chapter 2.3). According to Schmid (2015, 242-243), conventionalization "is defined as the continuous mutual coordination and matching of communica-tive knowledge and practices, subject to the exigencies of the entrenchment processes taking place in individual minds". In other words, then, it is the ac-commodation between idiolects (as carried out by their speakers), the levelling of variation between idiolects that ensures that speakers of the same speech community, i.e. speakers who interact frequently, produce mutually intelligible utterances. Schmid continues in saying that "[f]our types of conventionalization processes are distinguished in the model: innovation, coadaptation, diffusion, and normation". As part of conventionalization, they may be seen to describe various stages of a construction becoming conventionalized: it is innovated by one or multiple speakers, then adopted and adapted by other speakers, dif-fuses in the speech community and eventually becomes the norm. All these processes have happened, to some extent, to the constructions I will analyse below. But what happens if the process is disrupted or never fully comes to pass?

If we apply the notion of a feature pool to changes within the linguistic system, then it is equally important to look at innovations that do not spread. Since it is these innovations that are in competition with those innovations that eventually emerge as the preferred variant among the speech commu-nity, the factors that inhibited the unsuccessful innovations are the very same which would affect the successful ones, with the successful ones bearing some kind of advantage, be it inherent or caused by an external factor. Mufwene summarizes this notion as follows:

> There are several speakers whose peculiarities are not copied by others. One of the interesting aspects of language evolution is that there are a lot of innovations or deviations that do not spread within the language community. As a matter of fact, successful innovations or deviations, which are copied by other speakers, are quite few in number compared to numerous others that are produced daily and die immediately or are short-lived within a network of speakers. This state of affairs accounts for why communal languages are assumed to be generally stable, changing very little during (parts of) the lives of some speakers, especially with regard to grammar and phonology (consistent with Dixon's 1997 claim of punctuated equilibrium) (Mufwene 2008, 66).

A very similar position, with special regard to the history of Tok Pisin, has been taken by Mühlhäusler (1985, 118). He argues that its unsuccessful developments can tell us as much about the processes and factors of change and development as successful developments can:

> It is interesting to observe that in different varieties of Tok Pisin, pretty well all these possibilities are tried at one point or another. Only some of them get selected for use in the wider pidgin speaking community, however. The importance of mapping such unsuccessful developments together with the ones that eventually make the day cannot be overestimated.

Anyone agreeing will hopefully find the pages below enlightening in this regard. Before we delve into the analysis however, some issues and processes remain to be mentioned, among them the role of the size of a speech community.

2.4 Community size

Linguistic research has shown some interrelation between the size of the speech community and the structural properties of the language spoken in said community. Greenhill et al. (2018, 1-2) give an excellent summary of the various theories an arguments as to how language change and development on the one hand and the size of the speech community on the other could interact:

1. larger populations mean higher rates of change because they provide more opportunity for innovation
2. larger populations are less prone to random sampling effects that cause elements of language to be lost
3. larger populations have less stringent norm enforcement, benefitting change

4. larger populations have more robust transmission; more models to learn from increases success of transmission

5. exposure to more input allows more robust learning and retention of wider diversity

6. rates of change are faster in smaller population due to more rapid diffusion

7. language in smaller speech communities develop and retain more linguistic complexity

8. smaller speech communities show more diversity

9. smaller speech communities show more malleable linguistic representations, benefiting change

10. founder effects may accelerate language change in languages started in small populations, which increases loss of ancestral language elements

11. small populations are more heavily affected by language contact through trade and intermarriage across groups, accelerating rates of change

Of these claims, four are of high interest to the analysis at hand. First, the notion of faster change in smaller population due to more rapid diffusion aligns well with the evolutionary concepts borrowed from biology and incorporated into evolutionary linguistics. Just like in evolution, smaller populations may be affected by phenomena such as genetic drift, small speech communities may be affected by singular events which introduce new variants. Secondly, the notion that smaller speech communities allow for more complex structures and thirdly, that they show more diversity, are of special interest to language contact and the emergence of new languages, given that they usually arise on a small scale first. Finally, the idea that small populations are more heavily affected by language contact is of interest, as well. Together with the first notion, this suggests that the sociohistoric situation in which the variety at question may have first arisen needs to be examined carefully not only for possible linguistic influences, but for the circumstances under which they were exchanged as well.

To detail some of these notions, Lupyan and Dale (2010, 1), for instance, have shown that "languages spoken by large groups have simpler inflectional morphology than languages spoken by smaller groups as measured on a variety of factors such as case systems and complexity of conjugations". Wray and Grace (2007, 1) similarly found what they call exoteric communication systems tend to become simplified, while esoteric communication systems tend to have more 'complex' features. Nettle (2012, 1835) claims that "[l]anguages of small communities tend to have smaller phonological inventories, longer words and

greater morphological complexity than languages spoken in larger communities". The corollary thus seems to be an opposite tendency between the size of the speech community and the complexity of the language - at least on certain structural levels.

As Reali et al. (2018, 1) point out, this is not the case for certain non-structural features, since "an apparently opposite pattern appears to be observed in relation to non-structural properties of language: languages with large linguistic communities tend to have larger vocabularies of content words". They propose to solve this conundrum through computational simulation with a design based on a population of interacting agents in a network. Said agents were capable of passing linguistic conventions between themselves across links in the network, meaning that each agent could pass a convention to the neighbouring agents, but not directly to all agents in the network. Agents were capable of passing three types of conventions: those that they innovated spontaneously, those they had learned from others previously and those that they themselves had innovated previously. The concept behind such a setup was that conventions could be either difficult or easy to diffuse, depending on how many times an agent needed to be exposed to a convention to successfully acquire it, modelling 'real-life' language acquisition. As Reali et al. describe the setup, "in large groups, only linguistic conventions that are easy to learn, such as words, tend to proliferate, whereas small groups where everyone talks to everyone else allow for more complex conventions, like grammatical regularities, to be maintained". They sum up their results as follows:

> [T]he differential effects of population size on structural complexity and vocabulary size can be accommodated within a parsimonious model of cultural transmission constrained by one cognitive constraint: Ease of Learning. Linguistic innovations that are easy to learn tend to increase in number as a linguistic community grows, because the number of potential innovators increases, and innovations can spread more rapidly. By contrast, small linguistic communities favour linguistic innovations that are hard to learn because they require multiple interactions between individual speakers.

One additional factor, they allow, might be "the degree to which properties of language can be learned independently", pointing out that while structural features of a language often depend on each other (also see chapter 2.3.3 on the interconnectedness of structural features), vocabulary items can be acquired relatively independently of each other. This may affect the diffusion of certain structural features if part of the speech community lacks other structural features that the feature in question depends on. For instance, the fronting of a

certain syntactical constituent is more meaningful if the system as a whole has an established, somewhat more rigid constituent order.

Another recent study by Raviv et al. (2018) investigated the role of population size in an experimental set-up of two stages. In stage 1, six communities of four and six communities of eight participants were given the task of interacting using an artificial language they had to construct on the spot. Each community was given the same time to interact before a test round, meaning that both the smaller and larger communities had "the same amount of interaction and experience overall, but members of larger communities had less shared history with each other". In order to offset this difference, larger communities had additional communication and test rounds. In stage 2, the small communities received an equal amount of time as the larger groups. The results are as follows:

> [B]oth small and larger communities developed compositional structure over time (measured as the average correlation between labels' string distances and meaning distances in the community, [...]). Crucially, larger communities developed linguistic structure faster and more consistently than small groups. While there was no difference between the structure created by small and larger communities after eight rounds (seven communication rounds + test round), by the 16th round, larger communities had more compositional structure than small communities (Figure 1). In addition, small and larger communities showed similar trends of stabilization and conventionalization by the 16th round. Communicative success was not influenced by community size at any point in time. Finally, small communities showed significantly more variance than larger communities on all measures.

In line with the other studies above, Raviv et al. draw the conclusion that population size can affect the formation of linguistic structure. Larger communities, they surmise, develop structure more quickly and more consistently. Conversely, smaller communities are a more fertile ground for high diversity and more complex structures. If both studies discussed above can be taken as an indication for a strong tendency of complexity and variety decreasing as population size increases, this has important implications on the initial stages of the emergence of a contact language, especially pidgins and creoles. We can assume that for most pidgins and creoles, the community size under which they originate to be comparatively small. If historical data confirmed such a hypothesis, this would have three consequences.

First, we would see more variation in the emerging contact language than in its later stages (with increasing size of the speech community and increasing

opportunity for interaction reinforcing each other). This factor is compounded by at least two others. Emerging contact varieties are more likely to lack norm-enforcing structures. They will also lack structural means to express certain concepts, which may lead to parallel innovation - different strategies for expressing the same concept. We can therefore expect emerging contact varieties, especially those spoken in smaller speech communities, to show a higher degree of intra-group variation than later stages of the same or other languages.

Secondly, we may expect such varieties or stages of varieties with lower speaker numbers to be more structurally complex than later, 'stabilized' varieties. For emerging contact languages, this prediction has a caveat, however. Since many of the structural features may still be lacking during the innovation of a particular feature, or many features may be innovated in parallel, there is less of a structural scaffolding for new features to be based on. In other words, diffusion of certain innovations may, as described above, be hampered by the fact that necessary support from other structures is missing.

Finally, in terms of stabilization, there may be a certain, necessary threshold in the number of speakers before variation is levelled onto a more consistent language system. Naturally, this immediately raises the concern of where such a threshold should be set. Is the number of speakers necessary to create a viable creole really just one, as Bickerton (1990) has once suggested? What is the maximum threshold of variation at which we can speak of a 'stable' or 'stabilized' stage of a pidgin or creole? We shall return to all of these questions below when evaluating the various models of Tok Pisin's emergence. For now, we will turn to the question who it is within the speech community that the role of innovator falls to.

2.5 Who innovates?

In the development of pidgins and creoles (or, pidgins to creoles), a key point occurs when the language is transmitted to children for the first time. The switch from second or third or n-th language acquisition to first language acquisition has profound impact both on the future speakers and the structure of the language itself. In the context of creativity and innovation, it may seem logical at first glance for children to contribute a significant portion of innovative features to an emerging contact language. First language acquisition is, after all, full of structures and items that adult speakers of the same language would not use, whether they occur due to overgeneralization, incomplete acquisition or other factors. Furthermore, the adoption of a pidgin as a first

language means that its context of use will be greatly expanded, including situations in which its structures and vocabulary may not be sufficient for the task of successful communication yet.

As Senghas and Coppola (2001, 328) points out, "[s]ome linguists have proposed that child learners transform pidgins (simple systems developed by speakers of incompatible languages) into creoles (more complex languages that arise in later generations of such mixed-language communities)", quoting, among others, Anderson (1983); Bickerton (1984) and Sankoff and Laberge (1980). Their argument can be summarized thus: children take the impoverished or nascent structure of the pidgin and elaborate them, adding new constructions, items and structures in new usage contexts by applying what Bickerton would call the bioprogram.

Certainly, children possess some facility and ability to create linguistic structures. One extreme example where such abilities have shown themselves is the creation of Nicaraguan Sign Language (NSL), which Kegl and McWhorter (1997) as well as Goldin-Meadow (2004) attribute largely to children. An example from Senghas and Coppola (2001) will illustrate this matter. Among the features investigated was the use of spatial modulations. While similar between the groups surveyed - the older, first cohort, and the younger, second cohort - the second cohort made systematic changes to the language as they learned it. Senghas and Coppola points out that these changes were made on three levels: first, spatial modulations were more frequent in the second cohort; secondly, that they were "increasingly used for indicating shared reference" in the second cohort, which enabled long-distance grammatical relationships among words; and thirdly, that these changes accompanied an overall increase of fluency. In other words, the second cohort improved over the input they could have received, "taking a partially developed language and systematizing it in a specific way", despite having fewer years of exposure.

Senghas and Coppola argues against interpreting these results as mere evidence of regularization, as had been observed in other language acquisition studies, arguing that unlike "common" regularization, these changes left a lasting impact on the language and its future speakers, including adolescents. Neither did these children show error-correction or overgeneralization for the features observed. Quite to the contrary, the second cohort narrowed the function when compared to their input, with "the same sentence [having] fewer possible meanings" when using the patterns of the second cohort. Senghas and Coppola therefore conclude that "the second co- hort has reanalyzed the location of spatially modulated signs as indicating something akin to co-indexing" (Senghas and Coppola 2001, 327) and thereby increased the

specificity of Nicaraguan Sign Language. In his view, while the first co-hort did not have adequate time before reaching adulthood to fully develop it as a language, their systematization of the gestures and homesigns allowed the second cohort to have adequate material to build a full grammar.

As Mufwene (2008, 78) points out, however, "[t]he fact that, in the case of Nicaraguan Sign Language (NSL), the new "system" is largely a systematization of materials that were already available to the children in their respective native communities [...] reduces nothing of the important role that children played in elaborating a long-lasting communicative system". That role, however, was not one primarily of innovation, but of selection on a communal level. Note that selection on a communal level does decidedly not imply an elimination of inter-individual variation. In fact, as Senghas and Coppola (Senghas and Coppola 2001, 327) have, among others, shown, younger signers with sufficient exposure show greater systemacity, uniformness and fluency in their signs than adult signers do. Of the role of children in creating creole structures in particular, Mufwene (2008, 79) writes:

> Children did indeed play a non-negligible role in the development of these new vernaculars, but it was not that of creating a grammar where their parents would presumably have failed. It is not true that incipient pidgins have no grammars, although these are internally variable. Rather than creating new grammars for the overall community, children participated in the development of Creoles by selecting from the feature pools to which they were exposed particular subsets of features (including xenolectal ones) in manners that made their idiolects less different from each other than their nonnative parents'. Functioning thus as agents of normalization, they helped those features prevail over other alternatives, reducing the extent of variation observable in their xenolectal parents (DeGraff 1999a, 1999b). Creole children did this in the same way children everywhere normally contribute to changes and to divergence from their parents' language varieties, minimally in most cases except in exogenous settings, where the model speakers are the native speaking population if they have full access to it. Creole settings made variable access to native speakers of the colonial vernacular.

There is no conclusive evidence, he continues, that would indicate it is any more likely for creoles to derive their innovative structures from children than there is for European languages. As he points out, this is even more compelling for those 'innovations' we can trace back to some level of substrate influence. Since children who acquired the emerging creole as their L1 would have less reason to fully acquire the structures and patterns of the substrate language, there is a reduced chance they would introduce structures that so closely align

with the substrate. Furthermore, there is little argument to be made for strong parallels between child language and creoles beyond points such as the tired and disproven 'impoverished' morphological system. Mufwene, to this point:

> First, as argued by Slobin (2002), whether or not child language lacks inflections depends on what the target language is and how significant the role of inflections is in it. As Chaudenson (1992, 2001, 2003) points out, aside from the fact that the relevant European lexifiers are not so rich in inflections, they are also marked by inconsistencies that make them difficult for L2 learners to acquire. On the other hand, DeGraff (2001a, 2001b) shows that Creoles are not as deprived of inflections as has been claimed. In the very least, the old myth is not true of Haitian Creole, which displays some, both inherited from French and innovated thanks to patterns emerging from the new "system."

If children are not the primary innovators in creole emergence and in general language change, it logically follows that adults must be. This is not to discount the importance of children as first language acquirers in the process, however. Their role is to "help[...] determine which of those innovations become part of the communal language and which of the extant variants become recessive and may eventually die out of the ever-evolving language".

Placing the onus of innovation squarely - or at least, primarily - on the shoulders of adult speakers has at least four important implications for such innovations in contact languages. First, it means that substrate influence is available as a source for categorial and formal supply (notions that I will explore further in the following chapter). Secondly, the agency of speakers is generally higher for adult speakers than it is during first language acquisition. An adult speaker is, for instance, able to consciously violate structural rules or norms in order to express a novel concept. Consider, for instance, the fact that Tok Pisin has SVO structure for the most part. However, it is possible to front the object or patient of the sentence in order to focus it. Since Tok Pisin does not have a grammatical passive, violating the basic sentence structure in such a way allows for the expression of a similar concept. In first language acquisition, such a violation may be seen as an error, whereas after 'successful' acquisition, it can be a strategy. Thirdly, it means an increase of variation, which is more widespread among adults as described above. This can mean, as we will see several times below, that there is more than one viable strategy for expressing the same grammatical concept or closely related concepts. For instance, there are no less than five functionally identical complementizers in the Tok Pisin data for this thesis (see chapter 6.11). Fourth, and following from the third point, the fact that large-scale innovation happens in a speech

community composed mostly of adult L2 learners also means that regularization and standardization have not occurred on all levels of the language. On the one side, this enables variation, on the other side, it precludes the notion of acquisition errors, given that no accepted standard is established against which to measure what an 'error' is.

Whether children or adults are the primary innovators, does not, however, change the primary goal of linguistic innovation, which is, of course, successful communication, i.e. either successful encoding or decoding of a message. The difference then lies not between first and second language acquisition of an emerging creole, but between first language acquisition of a stable language system and first language acquisition of an unstable, emergent system. The processes are not distinct - acquisition happens on the same grounds in either case. The results are the same as well: a linguistic system is acquired. The difference is on the level of the language or the speech community itself in the latter case: whereas first language acquisition of a stable system affects the system itself only to a small extent, the selection of variants that children acquiring unstable systems engage in leads to some stabilization and regularization itself, eliminating certain unsuitable variants.

Naturally, which variants are available, and which variants may be desired, is the result of a complex equation covering linguistic supply and demand.

2.6 Linguistic supply and demand

The simplified picture of how pidgins, creoles and other contact languages emerge through language contact involves the source languages - superstrates, substrates and adstrates - providing the necessary grammatical and lexical input to create and shape emerging structures (Winford 2009, 22-24)[6]. In other words, both sides of the form-function equation stem from items, patterns and structures that exist within the source languages before they are introduced to the emerging language. An example is provided by the parts of the pronominal system of Jamaican Creole. Most of its elements can be traced back quite easily to that of its superstrate English, given that they are either identical or at least very similar in either form and function, or both, as shown in table 2.2 (adapted from Winford 2009, 323):

---

[6]Or, to put it more accurately, for the speakers of the emerging contact language to create and shape these structures based on the structures of the languages they employ in the contact situation.

Number person	Subject	Object
Sg, 1st person	mi	mi
Sg, 2nd person	yu	yu
Sg, 3rd person	im	im
Pl, 1st person	wi	wi
Pl, 2nd person	unu	unu
Pl, 3rd person	dem	dem

Table 2.2: The pronoun system of Jamaican Creole

There is, however, no requirement for both form and function of a linguistic item or pattern to stem from the same source, nor is there one for the demand and/or supply to be introduced by the same source. A distinction has to be drawn between *categorial* supply and demand on the one side and *formal* supply and demand on the other. The former refers to the need of a language system to express a certain grammatical category - such as past marking - (or the supply of such a category by another language system). The latter refers to the demand for a structural means of expressing said function (or the supply of a form which expresses the function). Demand in this case, is not an absolute force: it can be triggered (through societal changes or through language contact), but it does not have to be. There are, after all, plenty of languages whose speakers get by without certain grammatical categories. In other words, categorial demand can be weak or strong: if there is strong demand for a new function, then it is more likely that it is created. If there is weak demand among the speakers, then a change might not take place.

While it is possible that a source language provides both categorial and formal supply and demand, it is equally possible that contact with a source language only triggers structural demand, but the structural means are provided by another language or the material already existent in the emerging contact language itself, as we will see below.

The outcomes of individual contact situations differ greatly as to how much material and which items, patterns or constructions each source language provides, with one distinction in outcomes being drawn between radical, basilectal and intermediate creoles (Winford 2000, 214-215). Said distinction depends on how much the structures of the contact language still resemble their superstrates. This factor is, in turn, dependent on how much of its grammatical and lexical structure the emerging contact language takes from each of its super-, sub- and adstrates, respectively. The less material is taken from

the substrates, for instance, the more closely will those features not arising from internal innovation resemble the superstrate and vice versa. The balance of these contributions is determined by a number of factors, including whether a grammatical function is available in none, one or more of the source languages. Even when multiple sources are available, however, the lack of a certain grammatical feature in an emerging contact language (i.e., the strength of the 'demand' as described above) does not automatically lead to its direct adoption from a source language. Instead, such an emerging language may opt for one of three possible alternative strategies:

- Forego grammatical encoding of the function

- Encode the function grammatically by using different structural material (innovation by recombination)

- Create new grammatical structures to encode the function (complete innovation)

Referring back to the notion of categorial and formal supply/demand above, categorial demand in the emerging contact language arises from either contact with one or more of the source language(s) - which provides categorial supply - or from internal developments of the contact language itself. Once categorial demand is present, formal demand follows. It can either be discarded (option 1 above), filled by formal supply from the source language(s) (option 2 above) or by innovation in the emerging language (option 3 above).

A related concept has emerged in works of multilingualism. Riehl (2015, 108ff.), citing Matras (2009), for instance, differentiates between *matter borrowing* and *pattern replication*. The first relates to the transfer of linguistic matter, i.e. morphological or lexical material, from one language into the other, which would equate to the formal supply in the terminology above. Pattern replication, on the other hand, refers to the borrowing of more abstract structures from one language into the other, such as the use of pronouns in pro-drop-languages. Pattern replication might then be likened to categorial transfer, or categorial supply from one language into the other.

The notion of supply and demand does, of course, not explain how exactly that demand may be filled through linguistic processes. I would argue that one of the most important processes in this regard is grammaticalization.

2.7 Grammaticalization

2.7.1 General definition

Grammaticalization as a whole is a process understood to lead to the formation of grammatical items, constructions and patterns, i.e. linguistic elements devoid of lexical meaning. It is certainly not, however, a simple process, with many different subprocesses, mechanisms, factors and limitations involved, as the examples discussed below will show. For now, we shall content ourselves with giving a minimal definition of grammaticalization that would be able to reach a broad consensus.

This is by no means a simple endeavour. A review of the literature on grammaticalization, such as the seminal work by Traugott and König (1991, 189) indicates grammaticalization to be, for instance, a "dynamic, unidirectional historical process whereby lexical items in the course of time acquire a new status as grammatical, morpho-syntactic forms, and in the process come to code relations that either were not coded before or were coded differently". Relating back to the notion of solving a demand situation above, it is therefore the main way of either innovation by recombination. Lehmann (2004, 155), on the other hand, defines the "grammaticalization of a lexical sign" as a "process by which it loses its autonomy through increasingly being affected by the limitations of the linguistic system" [7], while Heine and Kuteva (2002, 2) define it "as the development from lexical to grammatical forms and from grammatical to even more grammatical forms". Far less complex definitions may also be found, as for instance in Croft (2006, 366), who understands it to simply be the "process, by which grammar is created".

The common core to all the definitions outlined above is that they understand grammaticalization to be a process, that is, an inherently diachronic phenomenon, through which grammatical elements are created. Traugott (2013, 201) remarks that "[w]hile most work on grammaticalization is diachronic, with focus on constraints of change, some is synchronic, with focus on 'principle[s] according to which subcategories of a given grammatical category may be ordered' [...]." The concept of grammaticalization being a process is thus a foundational one, and it may, in its effects, in turn be seen as a foundational part of language change. In how far it can be understood to be located on an exclusively or primarily historical or diachronic level, however, remains

---

[7]"Prozess durch welchen es seine Autonomität verliert, indem es zunehmend den Einschränkungen des sprachlichen Systems unterliegt".

unclear, not least because of the fuzzy borders between synchronic and diachronic linguistics in general - which is, of course, why the above quote has to be qualified by "most works". Surely the prototypical grammaticalization process is of a more diachronic than synchronic nature, but that does not preclude observations of grammaticalization categorically, though the latter would require the linguist to also answer how fast an element would have to be grammaticalized in order to qualify as synchronous grammaticalization. Put differently, we may be able, with sufficient synchronous data, to observe the onset of grammaticalization synchronously, but not the entire process; and we can certainly never fully claim a grammaticalization process to have stopped with only synchronous data.

The question of how grammaticalization relates to other processes of language change is especially relevant in the light of definitions such as the one by Croft quote above. By their vague nature, they run the risk of the process itself remaining vague and intangible as well as hard to differentiate from other processes of language change. Thus, Lehmann (2004, 155) already criticizes that such a definition would "necessarily rende[r] the concept wide and heterogeneous". This naturally leads to the question of which additional factors or properties have to be assigned to the process of grammaticalization in order for it to be delineated from the abstract process of language change. Based on the definition of Traugott and König, it may be prudent to first observe the initial input and results of the process in order to approach the concept of grammaticalization as a whole. Given a lexical unit as input, the output would be a grammatical, morphosyntactic form of some kind.

The idea of a lexical element as the initial onset point for grammaticalization serves as a critical foundation for many, if not all, definitions of grammaticalization. However, Heine and Kuteva, in the definition above, already admit the possibility of grammaticalization turning already grammatical items into even more grammatical items, i.e., that the initial input does not necessarily have to be a lexical element. Therefore, an adapted version of Croft's definition may be more suitable, in which grammaticalization is a process by which linguistic elements progress on a scale between less grammatical and more grammatical.

2.7.2 The grammaticalization process

Before taking a closer look at another part of some definitions of grammaticalization, namely its unidirectionality, it is prudent to take a closer look at the

concept of a process itself. One question would be which subprocesses grammaticalization consists of. Heine and Kuteva (2002, 2) list four mechanisms: desemantication, extension, decategorization and erosion. These describe, respectively, the loss of lexical content, the loss of usage limitations, the loss of categorical properties and the loss of phonetic material. Such subprocesses make grammaticalization a tangible phenomenon, which may be empirically observed. If we add the idea that grammaticalization occurs not only on purely lexical, but also less grammatical elements which can be processed into more strongly grammatical elements, this necessitates the concept of a grammaticalization scale or cline. Individual elements could then be located on such a scale. Parameters which could be applied in this regard have been put forward by Lehmann (2004, 154) and include weight, cohesion and variability on both paradigmatic and syntagmatic scales. If a linguistic element shows loss of weight and variability or increase of cohesion, it would be progressing towards the more grammatical element of the scale.

Another question would be what type of process grammaticalization is, and whether it may be approached through a categorization of the process. Traugott (2013, 270) indicates the different conceptions of grammaticalization as reduction on the one hand and expansion on the other. The first, which she also describes as "the 'traditional' or 'prototype' view of grammaticalization" is based on the idea that a lexical element exhibits richer content. Thus, the change into a grammatical element is accompanied by a change from a more complex to a less complex structure. The term reduction, then, relates to the area of complexity first, but also to the reduction of phonological and morphological material. Prototypical examples of grammaticalization phenomena therefore include languages rich in inflection, in which clitics and affixes are the main results. Opposed to that conception is the one of grammaticalization as expansion, which "[questions] the requirement of structural reduction and increased dependency" (Traugott 2013, 274). This approach is focused on the fact that resulting connectives or discourse markers can occur in syntactical contexts in which they were unable to occur in their original, purely lexical or less grammatical function. Their functional range has therefore changed (if they can no longer be used in the original function) or it has increased and expanded (if they can).

If grammaticalization can be understood as a process of both expansion and reduction, it follows that the categorization of the process is of no great help in searching for a minimal definition. The concept of unidirectionality also proves to be more complex than it appears at first glance. The issue is whether unidirectionality necessarily implies irreversibility, or if a process diametrically

opposed to grammaticalization, which has been dubbed degrammaticalization in some works, could construct less grammatical or purely lexical elements from more grammatical elements. But even if a process of degrammaticalization could not be proven, it does not automatically follow that grammaticalization always occurs along the same paths and in the same direction, as the analyses below will also show.

2.7.3 Sources and paths

While on the topic of paths and direction, it seems prudent at this point to define the notions of grammaticalization sources and grammaticalization paths. Two of the characteristics of some definitions of grammaticalization are its directionality and supposed irreversibility. The process itself implies some kind of progression on a scale from one concept - tied to a less grammatical or lexical item or construction - to another concept, which fulfils a more grammatical conceptual function. For instance, GO as the concept of spatial movement is processed to the concept of FUTURE, with, in English, the form *going to* progressing from lexical verb to grammatical marker.

The grammaticalization source (concept), in this case, would be GO or MOVEMENT. The target would be FUTURE. As such, the grammaticalization path leads from MOVEMENT to FUTURE. In more complex cases of grammaticalization, which involve multiple functions, these paths may become longer and more complex in turn. For instance, a form may, due to the concept it expresses, first grammaticalize to one target function, and then expand to others. Conversely, these paths may also involve grammaticalizations from different source concepts converging on the same target concepts, leading to competing forms for the same function. For instance, figure 2.3 is the grammaticalization map for the continuous and progressive aspects in Tok Pisin (see chapter 6.8.2). In this example, lexical items tied to the concepts of DO and STOP both grammaticalize to markers for the continuous aspect. However, they do so via different grammaticalization paths. DO progresses via the path of the progressive, whereas STOP takes the path through further lexical expansion to 'stay, remain' and the locative and existential.

While these grammaticalization paths are by no means universal to the world's languages, there are quite a few who have been observed in a number of crosslinguistic instances. A valiant effort in collecting and illustrating them is the *World Atlas of Grammaticalization* by Bernd Heine and Tania Kuteva. It lists hundreds of source-target pairs in 528 languages. Conceived of as a

```
progressive  ──  continuous  ──  existential
      |                               |
     'do'                          locative
                                      |
                                'stay, remain'
                                      |
                                   'stop'
```
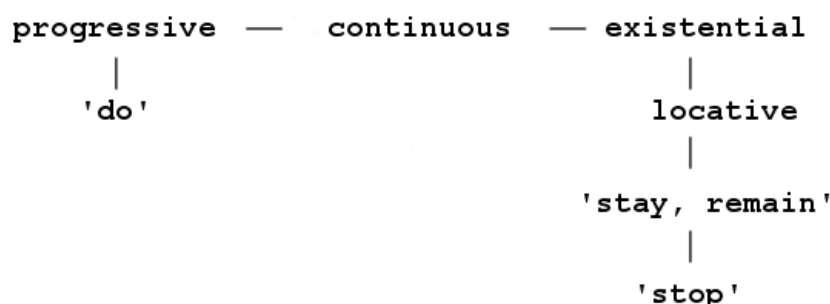
Figure 2.3: Grammaticalization paths for the continuous and progressive

reference work, the interested linguist will find, for instance, that PURPOSE frequently derives from the source concepts of ALLATIVE, BENEFACTIVE, COME TO, COMPLEMENTIZER, GIVE, GO TO or MATTER. Note that the source concepts have somewhat fuzzy boundaries and include both conceptual and more grammatical categories.

Some of the grammaticalization paths I establish for the constructions surveyed below are well attested in the *World Atlas of Grammaticalization* and other works. Others, however, are not. These are of prime interest not in that they occasionally run counter to the established paths - both in the opposite direction and on (seemingly) new paths. To adapt Robert Frost's famous words slightly: they took the path less travelled by. And that has made all the difference.

Once again, I stress that this is decidedly not to say that Tok Pisin and languages with similarly unusual structures employ different grammaticalization processes, nor are the outcomes entirely random. To the contrary, they employ the very same processes that other languages do, just with different input, different sociohistoric circumstances, and different output. Two very important parts of the innovation process as a whole are semantic and syntactic reanalysis.

2.7.4 Grammaticalization and semantic/syntactic reanalysis

If grammaticalization is the outcome of new structural material emerging in a language, then reanalysis is one possible pathway of achieveing that outcome. Langacker (1977, 58), in one of the most cited definitions, defines reanalysis in general "as change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifes-

69

tation". In other words, the form of the construction is unaffected, whereas its syntactic function changes. This change is often attributed to some inherent ambiguity of the construction due to either its semantic content or syntactic position, though this view has recently been challenged (see, for instance, De Smet (2009) or Detges (2002)).

An example of such a classical view of grammaticalization from Tok Pisin would be the reanalysis of *him* into the transitive verb marker (see chapter 6.4 below). Due to its position in the sentence, i.e., following transitive verbs (and, possibly, additional substrate influence), the pronoun eventually became a grammatical marker.

We might call this type of reanalysis syntactic reanalysis, as its main motivation seems to be the position of the construction in question within syntactic units. Aside from this type, there exists a second type of reanalysis which is based not on the syntactic level, but on the semantic level. Here, the adoption of a certain construction for another syntactic or functional purpose depends on some inherent semantic or conceptual quality the construction shares with the concept to be expressed. Here, we encounter a problem: if the speaker tweaks the functional properties of a construction in such a way, does she do so consciously? Is the ambiguity inherent to the construction, or created by the speaker?

Since speakers tend not to choose random words to serve new functions, I would argue that there is usually at least some inherent semantic connection between the forms and their new function. This semantic connection is what allows for semantic reanalysis: the change in the functional scope of a construction based on its semantic properties. This leaves us with two types of reanalysis: syntactic and semantic. In a more narrow sense, we might call them syntactically-motivated and semantically-motivated reanalysis, but for the sake of brevity, I will refer to them as syntactic and semantic reanalysis from here on.

The relation between grammaticalization and reanalysis is disputed. For instance, Newmeyer (1998, 238) sees it as an integral part of grammaticalization, arguing that "[t]he standard definition of grammaticalization incorporates the notion of reanalysis; no definition that does not do so seems particularly useful". Others, including Haspelmath (1998, 315ff.) argue that there is no relationship necessary for either process to function. Heine and Kuteva (2002, 5) thus conclude that "[w]hether grammaticalization involves reanalysis has turned out to be essentially a theory-dependent issue". For Hopper and Traugott (1993, 32), "reanalysis is the most important mechanism for grammaticalization". Similarly, Heine et al. (1991, 217) call reanalysis and

grammaticalization "inseparable twins". Detges and Waltereit (2002, 1) claim grammaticalization to be a speaker-based phenomenon, whereas reanalysis is a hearer-based procedure: for them, the speakers invent new discourse techniques to express their communicative needs, while the hearers reanalyse said innovations and eventually, through routinization, integrate them into language.

In the analysis below, I assume reanalysis to be one of several possible starting points for grammaticalization. Reanalysing a construction in question does not, in my view, automatically grammaticalize it. If we judge grammaticalization by Lehmann's parameters, the initial reanalysis lacks several of the characteristic features of grammaticalization. Furthermore, I agree with Detges and Waltereit that grammaticalization happens primarily on the part of the speaker, whereas the process of reanalysis is hearer-based (although in natural conversation, participants constantly switch between the roles of hearer and speaker and may thus engage in both processes). First of all, successful grammaticalization necessitates diffusion of the construction to hearers as well. Secondly, if ambiguity may be created through reanalysis, as has been suggested by them, it may be a more agentive phenomenon than previously assumed. If the speaker creates the ambiguity to express a concept he was unable to express before, this would suggest that there is an inherent speaker-based level to reanalysis.

Once a construction has been reanalyzed, it may be subject to all the further subprocesses of grammaticalization, including further expansion to other grammatical functions. It is not inconceivable either that reanalysis causes further reanalysis before a construction is fully grammaticalized. This is especially true in cases of contact-induced grammaticalization, which is the last theoretical concept to be introduced in this section.

2.7.5 Contact-induced grammaticalization

The notion of contact-induced grammaticalization is exactly what it says on the tin: it combines the field of language contact with the process of grammaticalization. Gast and van der Auwera (2012, 1) note the following:

> A more recent insight of contact linguistics, prominently put on the agenda by Heine and Kuteva (2003, 2005), is that language contact may not only lead to transfer or replication of matter or patterns, it can also trigger internal changes in a language under contact influence. The term "contact-induced grammaticalization" is now widely used for this process. Contact-induced grammat-

icalization is intimately related to, and in fact difficult to distinguish from, pattern replication.

In essence, contact-induced grammaticalization, they argue, combines pattern transfer with a following process of grammaticalization. They name the 'hot-news-perfect' of Irish English as an example, noting that it seems modelled on an equivalent construction in Irish. However, the Irish construction has not been borrowed directly, but rather has led to grammaticalization in English, "in so far as the newly created tense ("hot-news-perfect") has been integrated into the TAM-system of Irish English" (Gast and van der Auwera 2012, 2).

Heine and Kuteva (2003, 530) point out that "[g]rossly speaking, influence manifests itself most clearly in the transfer of linguistic material from one language to another, where linguistic material can be any of [several] kinds". These kinds include forms, meanings, form-meaning units, syntactic relations or any combination thereof. This means, of course, that a function can be transferred without the form itself being transferred, which is the idea behind contact-induced grammaticalization. Heine and Kuteva establish two subtypes of this grammaticalization, which they call "ordinary" and "replica grammaticalization". The second, which they argue is likely to be the more common, includes the transfer of a grammaticalization **process** rather than a concept being transferred itself. This is especially interesting in terms of language contact because it theoretically allows for processes to be transferred that have already run their course in the source language, which raises the question of whether the speakers need to be conscious of the process in order to replicate it in the target language. Heine and Kuteva (2003, 533, 539) summarize the processes of ordinary and replica grammaticalization as thus:

> Ordinary contact-induced grammaticalization
> a. Speakers of language **R** notice that in language **M** there is a grammatical category **Mx**.
> b. They develop an equivalent category **Rx**, using material available in their own language (**R**).
> c. To this end, they draw on universal strategies of grammaticalization, using construction **Ry** in order to develop **Rx**.
> d. They grammaticalize construction **Ry** to **Rx**.

As an example, Heine and Kuteva (2003, 534) put forward the French-based creole Tayo, which they claim has modeled its dual on the contact languages Drubéa and Cèmuhi. The marker in the form of a pronominal suffix *de* (Rx) was then deleveoped from material present in the language, namely from the

numeral for 'two' (Ry). Given that Drubéa and Cèmuhi have no numeral for 'two' as a dual marker, they have only been the source of the function, but not the form.

Replica grammaticalization
a. Speakers of language **R** notice that in language **M** there is a grammatical category **Mx**.
b. They develop an equivalent category **Rx**, using material available in their own language (**R**).
c. To this end, they replicate a grammaticalization process they assume to have taken place in language M, using an analogical formula of the kind [**My > Mx**] = [**Ry > Rx**].
d. They grammaticalize category **Ry** to **Rx**.

Examples for this second type are easy to identify, Heine and Kuteva (2003, 540) claim, "when the model language has developed a grammatical category by using a conceptual source that is rarely encountered crosslinguistically and where exactly the same source is used by speakers of the replica language". One of the examples they put forward for this second type is the hot-news perfect in Irish English (*She's after selling the boat*), which they trace back to the same construction in Irish. They see this as proof that speakers of Irish English have replicated the grammaticalization process that earlier took place in Irish.

Note that such a contact-induced grammaticalization process lines up very well with the notions of linguistic supply and demand as explored above. In the case of contact-induced replica grammaticalization, categorial supply triggers the categorial demand in the target language, which then solves its own formal demand by either copying processes from the source language (contact-induced replica grammaticalization) or relying on universal strategies of grammaticalization (ordinary contact-induced grammaticalization) rather than taking direct formal supply from it.

In those cases where formal supply is directly taken from the source languages, then, we usually see some form of reanalysis first, followed by later common (i.e. not contact-induced) grammaticalization. This may suggest that contact-induced grammaticalization presupposes a higher degree of agency by the speakers than pure grammaticalization does.
One structural difference that may lead to different processes than in 'ordinary' languages - whatever those may be - is that rudimentary pidgins lack systematical coding of categories such as tense, mood and aspect, compensating through intonation, context, etc. Arends and Bruyn (1994). One possible

solution to this problem is to grammaticalize content words into functional items and patterns (see chapter 2.7). Tense, mood and aspect markers are one area where this frequently occurs in contact languages (and "ordinary" languages as well). For contact languages, one causal factor may be that many of the common superstrate languages mark these concepts on the verb, and verbal inflection and auxiliary forms of the lexifier languages are frequently not transmitted into the contact language. Hence, TMA marking is one of the domains in which creation of morphosyntactic markers is frequent. Arends and Bruyn (1994) point out that the differences between grammar changes in what they call continuous languages and pidgins and creoles may lie in such developments. The need for communication in the latter leads to the forming of new categories (compare the notion of categorial demand in chapter 2.6) such as TMA markers. This is especially relevant in early stages. Once functional base needs are fulfilled, various factors behind grammaticalization more closely resemble continuous history languages (expression and creativity on the one hand and regularization and routinization on the other). In addition, contact languages may take "shortcuts through lexifiers or substratum languages" (compare the notions of formal and functional supply in chapter 2.6), which may help contribute to their supposedly rapid development. The problem, then, they point out, is to differentiate between lexifier and substrate language influence and internal grammaticalization, which is exactly what I will attempt to do in the analysis below.

Mufwene (2008, 173) places the origin of such grammaticalization (which he calls grammaticization) in "misidentification of forms and structures of the target language with those of the source language [which] leads to reanalysis or misinterpretation of the targeted forms and structures, i.e., to restructuring of the system". He likens the process to exaptations in biology or kludges in computing, suggesting that they are unplanned and occur only in the moment. To eventually become permanent, widespread changes to the structure of a language, they need to be spread and repeated by other speakers or be innovated across a great number of speakers in parallel. In coming up with these structures, they do not operate randomly but rely on grammaticalization paths that other, non-creole languages show as well (Mufwene 2008, 174):

> Studies of grammaticization based on non-creole languages have emphasized the importance of language-internal evolution (see, e.g., Hagege 2001). Their cross-linguistic comparisons have suggested "universal" paths of grammaticization, i.e., those that various languages, some of them genetically and/or typologically unrelated, have tended to follow, for instance, the tendency for PROGRESSIVE constructions to have developed from LOCATIVE ones, or for

TEMPORAL markers to have been extended from LOCATIVE ones. Among the things that the hitherto limited research on grammaticization in Creoles has revealed is that even these vernaculars follow more or less the same processes, with their creators making choices from among the structural options then available in the "lexifier." For instance, as shown in Mufwene (1996b), most Atlantic English Creoles have selected go as the FUTURE marker, owing apparently to the option of expressing FUTURE with *be going to* (or be gonna) + Infinitive in English. Interestingly, they have done it in different ways.

He further points out that some contact situations with English have led to different outcomes as well, drawing not just on English as their main lexifier, but incorporating material from the substrate languages involved "when there was enough critical mass among speakers of substrate languages with their particular alternatives" (Mufwene 2008, 175). The substrate languages in this case had no similar FUTURE construction with "go". He specifically names Tok Pisin as an "especially interesting" (ibidem) case among the contact situations, given that the grammaticalization of English *by and by* into the future marker from an adverb mirrors the construction of FUTURE in several Melanesian and Papuan languages considered substrates of Tok Pisin. In general, grammaticalization in creoles may be one of the processes that do not differ at the core but have different material and input to work with in contact situations.

In the end, however, it will be difficult to decide and eventually impossible to prove whether a grammaticalized item arose through contact-induced or normal grammaticalization due to universal processes, as the end result is the same. It is only through careful analysis of the sociohistoric contact situation, the various super-, sub- and adstrates involved and the grammaticalization paths of other, similar structures that we may construct a convincing argument for a particular origin. It is for this reason that the next chapter will give a detailed overview of the historical background of Tok Pisin.

3 Historical Overview of Tok Pisin

The following section will give a brief overview of the historical background against which Tok Pisin eventually emerged. As Smith (2002b, 13) notes, the question of 'the origin of Tok Pisin and other varieties of Melanesian Pidgin has generated considerable interest in recent years', which is a diplomatic way of stating the fact that we cannot, despite the extensive historical research done by Mühlhäusler and others, currently pin down the precise origin of Tok Pisin and its various predecessors. Part of the problem, of course, is where to locate the point at which Tok Pisin per se begins. Smith (2002b, 13) thus rightfully calls this date an "arbitrary delineation of this variety from a complex of English-based Pidgins which emerged in the context of the European exploration of the Pacific".

The first problem any attempt at determining a less arbitrary delineation of the origin of Tok Pisin faces is the difficulty of how far back into the Pacific contact situation one should look. Can the roots be found in the early exploration of the southwest Pacific by Spanish and Portugese explorers in the 16th century, which might be the ultimate source of "such words as *pikinini* (Portugese *peqeno*, 'small'), *save* (Portugese *sabe*, 'to know') [and] *kalabus* (Spanish 'calabazo', prison)" (Smith 2002b, 14), which are still used in modern Tok Pisin? Or do the actual roots lie with the beginning of the European settlement of Australia in the 1780s, which, as Baker (1993) has noted, many of the features of Melanesian Pidgin English can be traced back to? Alternatively, did Tok Pisin truly begin with whaling operations and the resulting contact and trade (Crowley 1990, 51)? Or did Tok Pisin not really start coming into existence until the plantations caused large-scale movement of labourers in the 1860s?

It seems likely that wherever one would place the marker, it would be at a point that, linguistically, was part of a continuing process that would eventually result in the emergence of Tok Pisin. What we can say with some certainty, at least, is that our interest should set in no farther back than around the 1850s. Smith (2002b, 13), too, points out that, "by far the greater part of Tok Pisin's history belongs to the last 150 years", i.e. from the 1850s on.

It begins, more or less, with what Smith (2002b, 14) calls a 'highly variable language with elements of English and various languages of the South and Central Pacific', often referred to as 'Pacific Pidgin English'.

3.1 A note on geography

Before we delve into the history of Tok Pisin proper, it is of use to recall the geographical area involved in its emergence. The areas of most interest to the present investigation are noted on the map in figure 3.2 (annotation mine). They are the eastern half of the island of New Guinea, called Papua New Guinea (1), the Bismarck Archipelago (2), the Solomon Islands (3), the Australian territory of Queensland (4), the New Hebrides (5, later called Vanuatu) and finally, the islands of Samoa (6). Most of the historical and social events that led to the emergence of Tok Pisin took place in this part of the Pacific. Since the map does not properly convey the distances involved, it is worthwhile to keep in mind that the distance, for instance, between Port Moresby on Papua New Guinea and the islands of Samoa is about 4460 kilometres, a distance spanning more than the entire continent of the United States. Even the distance between Port Vila on Vanuatu and Port Moresby still amounts to 2466 km. When using the term "Melanesia", it refers to the areas highlighted in the map below (Tintazul 2014):

Figure 3.1: The geographical extent of Melanesia

Figure 3.2: Map of the Pacific area of interest

3.2 Early contact (before 1860)

As mentioned above, it is impossible to pin down Tok Pisin's exact origin. We can be reasonably sure, however, that one of its predecessors is to be found in the Pacific Jargon English which emerged during the first contact situations between Europeans and Pacific Islanders. While, according to Mühlhäusler (1985, 37), "the area of New Guinea and the islands of the Bismarck Sea feature very marginally in these contacts", Smith (2002b, 13) asserts that even "items of vocabulary or grammar once seemingly pinned down to a definite origin have repeatedly been shown to have been attested in earlier periods", suggesting that there is merit in looking even beyond the date of 1860 in determining early source structures on which Pacific Jargon English might have been built. Of most interest, however, are the three "main waves", as Mühlhäusler (1985, 37) calls them, beginning with the whaling period during the 18th century, followed by the sandalwood trade during the 1830s and the trepan trade in the 1840 and 1850s. It is these waves which must have brought the Europeans and Pacific Islanders into ever closer and ever longer contact, since the latter trade activities would have required longer shore stays than the former (Churchill 1911, 7).

And it is trade that would have been of major interest to both parties involved in these contact situations, not missionary work, nor yet the establishment of plantations or colonies. Mühlhäusler (1985, 38) cites distrust on both sides – the Europeans taking the natives for treacherous cannibals, while the Pacific Islanders remembered earlier, hostile contacts – as a further factor in limiting interactions solely to trade. Of more interest to linguistic endeavours would be the contact situation not directly on the trading posts on the coast, but on the ships themselves, which would take on indigenous people as crew. Mühlhäusler (1985, 38) cites Reinecke (1937, 534-535) in saying that these new crewmembers "communicated with the crew in a lingo consisting of broken English and words from their own language" and points out that speakers of various pidgins would have met on the ships. As for how much of an impact these contact situations had on the later development of the various Melanesian Pidgins, we cannot be entirely sure. The numbers involved were very likely to have been very small, and there is considerable doubt as to whether these situations resulted in a stable pidgin at all, or rather only in what Mühlhäusler (1985, 38) calls "a number of unstable varieties of Jargon English in various parts of the Pacific Ocean". Let us take a look, however, at the preceding activities, which started with whaling.

3.2.1 Whaling

As Lever (1964, 33) notes, 'serious whaling [in the Pacific] did not begin till the voyage in 1789 (...)'. The peak of whaling operations in the area was reached by the 1840s, with decline setting in after 1860s (Mayr and Diamond 2001, 30). The vast majority of whaling vessels were sailing under an American flag. Contact with the natives was of a limited nature, as Mayr and Diamond (2001, 30) report:

> Whalers occasionally landed, but more often traded with islanders in canoes that came out to the ships. The whalers sought fresh food, water, sex, tortoise shell, and curios, in return for which they gave iron tools (axes and machetes), nails, cloth, and glass bottles.

In terms of a timeframe, Tryon 2004, 111, quoting Bennett (1987), suggests that, for instance, on the Solomon Islands, "the first whalers in the Solomons began to appear in the first decade of the nineteenth century, their numbers increasing by the early 1820s and reaching their peak in the 1840s and 1850s." While it is impossible to know the exact numbers involved in the contact situation, there are some records that allow us to take an educated guess. As a rough guide, Bennett (1987, 350), summarized in Mayr and Diamond (2001, 30), gives the following numbers for whaling contacts and sightings between 1799 and 1887 on the Solomon Islands:

Island	Shtl[1]	Ch	Simbo/Gizo	NGa	Ys	Russells	Fla
Number of whalers	14	13	7	1	6	1	2
Island	Guad	Mal	SCr	Ren	OJ	Sik	-
Number of whalers	5	14	39	5	7	15	-

Table 3.1: Number of contacts between whalers and the native population on the Solomon Islands, 1799 to 1887

While these numbers indicate that whaling and related contact did occur in the area that is of interest to us, the numbers hardly seem sufficient to talk of an extended contact situation. As Tryon and Charpentier (2004, 112) note, only "a few [Solomon Islanders] worked as crewmen aboard whaling ships", though it was they that would later "act as guides and interpreters for visiting

---

[1]The abbreviations refer to the names of the islands as follows. Shtl: Shortlands; Ch: Choiseul; NGa: New Georgia; Ys: Ysabel; Fla: Florida; Guad: Guadalcanal; Mal: Malaita; SCr: San Cristobal; Ren: Rennell; OJ: Ontong Java; Sik: Sikaiana.

Europeans". Contact was further complicated by incidents like a particularly memorable one in 1860, when whalers were killed at Mono and Makira Harbour on the Solomon Islands (Bennett 1987, 32). Even for those natives brought onboard, the employment situation would not have been stable, as whaling ships were notorious for frequent rotation of their crews.

3.2.2 Trade

Following (and partly overlapping) the period of whaling, the second wave of European contact with Melanesia centered around the trade of two goods of particular interest to China: sandalwood, which was burned as incense, and an edible sea slug called bêche-de-mer. It is from the latter that today's Bislama received its name, indicating the significance of the trading situation for linguistic contact.

Interest in the sandalwood trade began to spike in the Melanesian region around 1841, after a significant amount was discovered near New Caledonia. At the time, sandalwood was coveted by the Chinese. This in turn made it especially valuable to British colonies in Australia who were able to sell to the Chinese. Crowley (1990, 52) notes that sandalwood provided colony ships with a cargo on the outward journey to China, which were necessary to load tea. By selling sandalwood to the Chinese, they were able to significantly reduce the cost of the trips. The discovery of these reserves, then, was what formed the 'fourth and final sandalwood episode in the Pacific Islands' (Shineberg 2014, 8) and finally provided incentive enough for Europeans and colonial merchants from Australia to engage with the island groups of Melanesia, which "were almost unknown to the European world in the early nineteenth century" (Shineberg 2014, 9) for a number of reasons:

> 'Their warriors had a reputation for ferocity and for cannibalism. The extraordinary multiplicity of languages increased the difficulties of contact by enhancing the possibilities - already great - of misunderstanding. Except in New Caledonia and the Loyalties the political units were smaller than in Polynesia and the rules less powerful, so that it was harder for a foreigner to gain safety by winning the support of a chief. The islands acquired a bad reputation for 'the ague and fever and although the New Caledonian group was not in fact malarial, it was probably not known to be healthier than the rest. And Melanesian women, on the authority of Cook himself, were generally less hospitable than their Polynesia sisters. In fact to make a deliberate voyage to this area worth while, a strong incentive was required. [...] This was provided

by the discovery of sandalwood in this region. Sandalwood traders pioneered this area, mapping the coasts, noting winds, current and reefs, in many cases the first Europeans to land on some shores, and certainly the first class of Europeans to appear in numbers and at regular intervals.'

According to Clark (1979, 36), exploitation of the sandalwood resource was completed only 25 years later. While it was happening, however, it required longer stays on shore for the traders than whaling did. It was during that period that "large and permanent shore stations were established [... and that] Melanesians from a variety of these islands began to work on the sandalwood ships as crew alongside the previously exclusively European and Polynesian crews on these ships" (Crowley 1990, 60f.). In addition, a significant number of the workers processing the sandalwood for trade was Melanesian in origin (Romaine 1992a, 34f.). By the early 1850s, when the sandalwood trees had become more rare due to the exploitative harvesting the Europeans were promoting, Melanesian labourers would also be involved in their next ventures: trading the aforementioned sea slugs as well as turtle shell, copra, leather, mother-of-pearl and kauri resin. With sandalwood and sea slugs their major exports, shore stations in the area continued to operate until around 1865 (Crowley 1990, 62). The extent of these trading operations is extensively illustrated by Shineberg (2014, App. 1), who records more than two hundred individual trading voyages just between the years 1841 and 1855. However, the sandalwood trade increased not only the contact between Europeans and Melanesians on the shore stations, but on the vessels themselves as well. Crowley (1990, 64) notes that by "the early 1850s, the number of Melanesian labourers on ships increased dramatically [and that] by the mid- to late 1850s, ships' crews in the sandalwood and sea-slug trades consisted predominantly of Melanesians, with just a small proportion of Europeans and Polynesians".

3.3 German New Guinea and British New Guinea

As it had happened with whaling before, the slowly accelerating decline of one commercial activity did not mean the end of European engagement in the region, but merely a reconcentration of efforts in other types of commerce. Starting in 1863, Melanesian labourers were recruited to work on overseas plantations mostly located in Queensland, New Caledonia, Fiji and Samoa (Tryon and Charpentier 2004, 173). Due to the American Civil War and associated blockades, prices for cotton were sykrocketing, leading to the establishment

of cotton plantations first in Queensland (ca. 1863) and on Fiji (early 1860s) and later (1867) on Vanuatu as well. All of these plantations faced a severage shortage in labour force, giving rise to what can only very euphemistically be called a "massive recruiting campaign, often referred to as 'black-birding', for between 1863 and 1907 more than 100,000 Pacific islanders [...] were recruited, often against their will, for contract periods of three years" (Tryon 2004, 173f., cf. Shineberg 1999). Since the work involved was physical labour, the vast majority of recruits tended to be male. According to Tyron (2004, 182), "figures for the period show that female labour constituted only 8 per cent of the Pacific Islander plantation labour force". The numbers of 'recruits' differed wildly between the various islands and the mainland of Papua New Guinea. Table 3.2 below, adapted from (Tryon and Charpentier 2004, 176f.), illustrates this:

Year	Loyalties	Vanuatu	Solomons	PNG	Kiribati	Other
1863	-	67	-	-	-	-
1864	-	134	-	-	-	-
1865	-	148	-	-	-	-
1866	36	141	-	-	-	-
1867	329	874	-	-	-	34
1868	280	625	-	-	-	33
1869	-	313	-	-	-	-
1870	27	607	-	-	-	9
1871	292	978	82	-	-	-
1872	44	416	-	-	-	-
1873	7	987	-	-	-	-
1874	47	1332	124	-	-	-
1875	5	1931	728	-	-	17
1876	-	1575	74	-	-	39
1877	-	1986	-	-	-	-
1878	-	1218	240	-	-	5
1879	-	1821	354	-	-	7
1880	-	1934	61	-	-	-
1881	-	1976	641	-	-	-
18821	-	2699	440	-	-	-
1883	-	2877	1127	1269	-	-
1884	-	1010	714	1540	-	-
1885	-	1379	533	-	-	4
1886	-	1148	444	-	-	3

1887	-	1431	553	-	-	4
1888	-	1125	1143	-	-	23
1889	-	1412	620	-	-	-
1890	-	1294	1165	-	-	-
1891	-	534	516	-	-	-
1892	-	229	235	-	-	-
1893	-	714	416	-	-	-
1894	-	806	945	-	-	-
1895	-	519	577	-	-	-
1896	-	359	423	-	-	-
1897	-	201	733	-	-	-
1898	-	455	721	-	-	-
1899	-	674	848	-	-	-
1900	-	859	884	-	-	-
1901	-	530	1151	-	-	-
1902	-	264	875	-	-	-
1903	-	374	663	-	-	-
1904	-	19	59	-	-	-
1905	-	-	-	-	-	-
1906	-	-	-	-	-	-

Table 3.2: Number of recruited Queensland plantation workers and origin

As the table above shows, the majority of the labour force employed on the plantations in Queesnland came from Vanuatu and the Solomon Islands. Similar numbers can be found for the plantations on Fiji, where 52.06 per cent of the workers came from Vanuatu at first, before Solomon Islanders made up the vast majority of workers recruited (Tryon 2004, 182 citing Siegel (2009)). On the other hand, there were almost no workers recruited from the Papua New Guinea area, and those who were tended to come from the surrounding islands rather than the mainland. The reason for this occurred in 1884, when the *Deutsche Neuguinea-Compagnie* flew the German flag over north-eastern New Guinea and the Bismarck Archipelago as well as parts of the Solomon Islands. In the same year, the southeastern part of the mainland was officially made a British protectorate, to which adjacent islands were added in 1888 to form British New Guinea, which would be placed under control of the Commonwealth of Australia in 1906. The German government eventually assumed direct control over the northern part in 1899, making it part of the larger

colony of German New Guinea, or *Kaiser-Wilhelmsland* Sack (1972).

Mühlhäusler (1985, 44) goes so far as to say that "one could almost call it the year of [Tok Pisin's] birth". It certainly had far-reaching immediate consequences. The labour trade with most plantations in the Pacific Area was stopped by order of the new German administration. Only the German-owned plantations in Samoa would continue to receive workers from German New Guinea.

In Mühlhäusler's view, it was the plantations who served as a catalyst for both the stabilization and nativization of Tok Pisin. One of the texts Mühlhäusler at al. (2003, 140f.) provide in their collection relates the tale of 'How Tok Pisin came to Tumam' during the plantation period. An excerpt will suffice here to illustrate the situation[2]. Of the speaker, Romaine (1992a, 39) says that he "is a middle aged man', part of a generation with 'good knowledge of the language".

> 'Brata bilong mi mipela ol i long stesin. Na mipela i no save long Tok Pisin tu. Brata bilong mipela ol i go long stesin, orait ol i kisim save long stesin. Ol i kam bek orait ol i tok pisin. Na mipela save mipela i save samting i klia longen mipela i save. Tasol samting i no klia long em i hat liklik orait mipela mas haskim ol. Tok: 'Dispela samting kolim olsem wanem?' Orait, ol i tok: 'dispela samting em Tok Pisin ol i kolim olsem.' Orait i go i go i go i go. Woa i kamap, orait mipela i klia gut long tok pisin. Sampela lapun man long ples, mipela ol manki mipela i kisim save long tok pisin. Woa i pinis mipela i kisim save nau. Planti masta ol i kam insait long mipela, kiap o kampani masta ol kam bek, ol i wokim tret stua, o mipela i gat bisnis, na mipela olgeta meri man mipela i gat liklik save long tok pisin na ritrait nau. Na bipo, misin tasol i kam, wanwan man tasol i kisim save long Tok Pisin long misin. Taim misin i kam long ples. Orait, mipela i kisim save long tok pisin. Planti manmeri o mikpela pikinini ol pikinini mipela i no save Tok Pisin. Ol i tok olsem wanem, mipela i ting i hat, tasol nau, mipela i klia gut long Tok Pisin. Nau mipela i klia gut nau, husat narapela ples i laik kam, mipela tok susta, brata, kandare [...]'[3]

---

[2]Note that this tale was collected as a later recollection, and does not reflect the language as spoken at this time.

[3]Regarding Tok Pisin, me and my fellow villagers went to a station/plantation and we did not know Tok Pisin. Our relatives went to a station and they learnt it there. When they returned they spoke Tok Pisin. Some Tok Pisin expressions were clear to us but we were not so clear about others and we had to ask others: 'what do you call this thing?' They answered: 'in Tok Pisin it is called this or that.' Well, time passed, the war came, and we were pretty good at Tok Pisin. Some old men in the village - we were young boys then - we acquired a knowledge of Tok Pisin. The war came to an end and we learned. Lots of Europeans came to us, the patrol officers and business men returned, they opened up trade stores and started businesses and we had some knowledge of Tok Pisin and could read and write. In earlier times, only missionares came and only a few

The beginning of German New Guinea also saw an increase of missionary efforts. While in other German territories in the Pacific, Christianity had already made inroads, missionary efforts in German New Guinea "were still in the first flush of contact, wrestling with the frontier problems of rejection and acceptance as an alien force" (Hempenstall 1975, 47). Faced with the geographical, sociological and linguistic diversity of the region, as well as competition between the various Catholic and Lutheran missions themselves, it would be years before they found acceptance with the native population and success in their missionary goals. However, they eventually "played a significant role in helping to resolve conflict, both between islanders themselves and between them and the European community" (Hempenstall 1975, 47), allowing the German administration to "capitaliz[e] on mission success in promoting trust and security to widen its sphere of influence". While the missionaries' conversion efforts differed greatly in their success, mission society became an important part of interaction between the Europeans and the local community:

> [...] mission society was very often the first to be accepted as an institutional ally of the indigenous community in the eyes of the people, competing with business and government for its loyalty and resources. Mission societies could often command a legitimacy and assert an authority over islanders which was more efficacious than that of the secular government. [...] Numbers had something to do with it in New Guinea: missionaries comprised 25% of the European population throughout the period of German rule and were in more steady and familiar contact with the local villagers than the district office-dwelling administrator. (Hempenstall 1975, 50)

Naturally, mission efforts also included teaching the native population. As Kempkes (1968, 67) notes, "[o]ne of the first tasks of missionaries is always to establish the Church and teach the gospel, [which is] why the first centers of teaching in a mission country like New Guinea were always Catechetieal training centers". However, the language barrier in German new Guinea proved to be a significant obstacle in this regard, especially when it came to spreading a mission's influence beyond one station and the local tribes. As Mihalic (1977, 654) comments, no tribe would accept the other tribes' language over their own, the languages were structurally very different and if you had to use a foreign language, why not opt for English or German in the first place? Instead,

---

people acquired a knowledge of Tok Pisin at the mission. When the mission came to the village we acquired a knowledge of Tok Pisin. Many people and children did not know Tok Pisin when we were children. How did they talk? They thought it was difficult but now we know Tok Pisin well. Now we know it well and whoever wants to visit us from another village we call them sisters, brothers and uncles.

missionaries came to increasingly rely on Pidgin for their communication with the islanders, a fact that would be recognized by missionary administration in the Inter-War Period as well. The German administration, too, had its difficulties with the language situation, and was less eager to embrace using Pidgin. In the beginning of the German administration, official policy was to teach natives the German language and propagate it across the territory. Missions which taught in German received subsidies, but as described above, these were not enough to overcome the structural difficulties of teaching in German (Johnson 1977, 429). Apparently "horrified at the possibility of this 'horrible jargon' even becoming recognised as a language" , the German administration even attempted - unsuccessfully - to promote Malay as the official language (Mihalic 1977, 653f., citing Höltker (1945)).

Not only did the missions themselves serve to establish peaceful relations between both Europeans and natives and between various native tribes themselves, the German administration also had a vested interest in ending inter-tribal warfare. In what some have called a 'pax germanica', their efforts included "the gradual pacification of New Guinea, the termination of inter-tribal warfare and the expansion of effective government control over wide areas" (Wurm and Mühlhäusler 1985, 48). This enabled cross-tribal communication. In addition, village administration fell to tultuls (interpreters), which usually were returning plantation workers who had learned Tok Pisin and were able to communicate with administration officials. This new position of power within a village brought a new prestige to Tok Pisin, increasing the eagerness of young men to acquire it in order to climb the social ladder. They were further incentivized - or forced, depending on how one looks at it - to "enter the money economy" (Romaine 1992a, 38) of the Europeans by the introduction of the head tax and the goods sold in European-owned stores. While most of the population had survived through subsistence farming to that point, tobacco and liquor were mostly available through money. And even for those that were not lured by such goods, the head tax required them to work in order to meet their payments.

Away from the plantations and the servant-master system entrenched there, the varieties of Tok Pisin which emerged in the homeland more and more became means of intertribal communication instead of just being associated with the Europeans. Advances in infrastructure - roads, patrols and expeditions - further served to increase regional mobility along with the missionary efforts. While Tok Pisin may have become standardized on the plantations, it was back on German New Guinea and the islands of the Bismarck Archipelago that it was nativized (Romaine 1992a, 38).

3.4 World War I and the Inter-War Period

Within the first year of World War I, Australian troops occupied German New Guinea at the request of Britain. There was only a token German force stationed in the colony itself and the only naval force was the vastly outnumbered German East Asia Squadron, which was spread out over the islands on routine missions. German resistance thus proved only nominal. There was really only one battle of consequence being fought over the radio station at Bita Paka on the Bismarck Archipelago. On September 11th of 1914, the Australian Naval and Military Expeditionary Force landed on Kaiser-Wilhemsland and assumed control. The Treaty of Versailles would formally place the entirety of northeastern Papua New Guinea under Australian administration as a League of Nations Mandated Territory known as the Territory of New Guinea Hudson (1965).

The Inter-War period also saw an increasing influence of the missions on Tok Pisin itself. While World War I ended German administration of New Guinea, the German missionaries continued their work and "continued to have a great deal of linguistic influence" (Romaine 1992a, 44). The three most important Catholic missions at the time, the Rabaul mission centered on New Ireland, the Divine Word Missionaries and the Alexishafen missions on the mainland, began to use Tok Pisin in their missionary efforts in earnest. The former produced an array of both grammatical treatises and religious material in the language, and the latter made Melanesian Pidgin its official means of communication in 1931, having already produced and used educational material in all their schools (Mihalic 1977, 654f.). The missionaries at Alexishafen followed suit in 1939. It is worth noting that while the missions are often credited with some measure of standardization they brought to Tok Pisin, especially as a written variety, they did not immediately agree on either orthographical or functional variants. Mühlhäusler et al. (2003, 69) give the following three different beginnings to the Lord's Prayer:

> Alexishafen.
>
> Fader bilong mifelo, yu stop long heven. Ol i santuim nem bilongyu. Kingdom bilong yu i kam. Ol i hirim tok bilong yu longgraund olsem long heven. [...]
>
> Vunapope.
>
> Papa bolong mipela i stap antap, naim bolong ju i tambu, lotu bolongju i kam, mipela daun olosem ol antap i harim tok bolongju [...]
>
> Rabaul.

> Papa bilog mifela, iu stop an top alog [2] peles bilog iu, i qud mifela sigsig out tru alog nem bilog iu; i moa beta ol a fasin bilog iu i stop oltuqeta peles. I qud mifela mekim tru ol a lo bilog iu, ol a sem oltuqeta man i savi mekim alog peles bilog iu. [...]

The Australian administration, on the other hand, seemed to have little interest in interfering with the missions' schooling, and not much more in operating their own schools. Kempkes (1968, 68) reports that "at the outbreak of World War II the administration conducted only a very small number of schools. Five of its eight schools were teaching natives." The Catholic and Lutheran missions, on the other hand, operated close to 3,000 schools according to his numbers (1968, 68).

3.5 World War II

3.5.1 Historical developments

As it did for so many other countries, World War II brought about rapid and irreversible change in almost every aspect of life on Papua New Guinea. Mühlhäusler (1985, 55) starts the respective chapter in the *Handbook of Tok Pisin* with the following quote by Ryan (1972, 121):

> For New Guinea itself the war was the most cataclysmic event in the country's whole history. Between December 1941 when the Japanese struck and August 1945 when they surrendered, changes occurred or were set in motion which far exceeded in their effects the original coming of the white man (which had been local and gradual) or the results of any natural catastrophe of disease or volcanic activity.

The New Guinea campaign of the larger Pacific War theater began in January 1942. The Empire of Japan initiated aggression by invading first the Territory of New Guinea (the former German colony) in January and Papua (the former British protectorate), both of them now under Australian control, in March. Western New Guinea, administered by the Netherlands, was occupied in March as well. However, after this first somewhat successful push, the second phase of the Campaign resulted in heavy losses and eventual defeat for the Japanese. Australia turned to the United States for help, and after an assault on the strategic port of Port Moresby had been rebuffed by combined Australian and US forces and the Japanese had lost strategic initiative, the US navy

established effective blockades, denying crucial supplies to Japanese garrisons. The remainder of the Campaign would be largely fought on land, with naval support being crucial for ensuring Allied supply lines stayed open and Japanese supply lines remained cut (Stevens 2001).

The following years, and particularly the Allied offensives of 1943 and 1944 were not only "the single largest series of connected operations Australia has ever mounted" and "saw Australians fighting some of the hardest battles of the war" against "a determined and often desperate enemy" (Stanley 2003), they also led to much closer contact between Australian forces and the native population:

> A different sort of white man was seen for the first time in the Australian soldiers whose humanity, informality and willingness to labour in the sun and the mud were in startling contrast to the rigidity and aloofness of many of the pre-war white residents. (Ryan 1972, 223)

Both the Japanese, in their role as the invading force, as well as the Allied forces in their supposed role as protectors, quickly realized that they depended on the native population and their goodwill for success. Therefore, both sides engaged in massive propaganda efforts, including millions of leaflets which were dropped onto the Melanesian islands. The content and purpose of those leaflets war manifold: dehumanising the other side, instructing to help the own side, warning of impending bombing raids, instructions to help captured or shipwrecked soldiers, emphasizing victories, belittling the military power of the opposing force, and so on (Clark 1955, 11f.). One exemplary such leaflet preserved at the Monash University Library (2009) contains the following text next to two illustrations showing a native finding a wounded Japanese soldiers and leading Australian troops to him:

> TOK BILOG GAVMAN
> Sipos yu painim sompela Japan i les long pait, yu gifim dispela pas. Sipos i savi wakabaut, i kan kam ontaim yupela nau painim soldia bilog yumi. Im i sik tumas, orait, yu brinim tok.
> Tok im ol gut, mipela nokan kilim ol, kalabus dasol, nau salim ol iki long Astralia, na weitim pait i pinis.
> WOK BILOG GAVMAN. I GAT PEI.

In essence, the leaflet instructs a native who finds a Japanese soldier who can no longer fight to give the leaflet to the soldier. The native is to either lead the Japanese soldier to Australian troops or, if the soldier can not move, to let

Australian forces know his location. It also promises that Australians will not be able to kill him, but merely imprison him until the fight is over.

3.5.2 Linguistic developments

World War II constituted a turning point in the development for Tok Pisin in multiple ways. First, European interest in the language and its use increased sharply. Secondly, new functions and functional elements were added as it was used in a variety of new contexts. Thirdly, the resulting use by European forces spread the language further, geographically, than it had ever been carried before. Fourthly, the leaflet campaign served to encode a written variety of the language far beyond what earlier bible translations had done. Each of these was significant in their way.

As regards European interest in the language, whereas earlier interest among Europeans was limited to the works of missionaries, its potential use in the war efforts set off a suddenly heightened interest in the language. It is from this period that the first comprehensive grammars of the languages stem, of which "by far the most valuable contribution is [Robert] Hall['s]" *Melanesian Pidgin English* (Wurm and Mühlhäusler 1985, 55). It was this book that the US forces based field material for its soldiers on. The Japanese, too, tried to learn Pidgin English from Papua New Guineans (Lawrence 1964, 107). This had both immediate and long-lasting effects on the language itself, which brings us to the second effect. Mühlhäusler (1985, 56), for instance, reports:

> The most important aspect of the interest shown in Tok Pisin by the forces fighting in New Guinea was the development of two new functions for this language, that of promoting solidarity between the occupying armies and the indigenous population and that of large-scale social control through Tok Pisin media.
>
> The breaking down of the prewar social barriers and the development of a new solidarity between Australians, Americans and New Guineans finds its linguistic expression in the use of the inclusive first person plural marker yumi to refer to whites and blacks alike ("soldia bilong yumi i banisim ol Japan" - pamphlet dropped by the Australian Army near Dagua - East Sepik District). This constituted a significant change from the prewar masta - boi pattern.

Third, as mentioned above, both the leaflet campaign and land-based military efforts served to spread Tok Pisin in different ways, building on earlier missionary efforts. According to Clarke (1955, 12), "almost invariably there would be mission-trained "tultuls" or interpreters in each village who could read Pidgin

and convey the messages to others". He quotes the report of one mission, which confirms that "Second plane saw natives reading pamphlets dropped by first plane, gesticulating wildly" (1955, 12). Many others were involved in closer contact with European forces. Mühlhäusler (1985, 58) reports the following regarding the recruitment of the local population:

> To overcome the transport problem, tens of thousands of males over the 'apparent' age of 14 were conscripted by the Australian and American forces to serve as labourers and carriers . At the peak of the war activities their number was 55,000. Recruitment took place in Tok Pisin-speaking areas as well as in remote areas.

He further quotes Vader (1972, 35) in saying that "Australians, Americans and Japanese learned Pidgin - and so did the wild-looking tribesmen recruited from remote mountain ridges and valleys". This geographical and demographic extension of Tok Pisin, along with the large-scale population movements the war effort enabled and which continued after the war, "is certain to have accelerated the neutralisation of regional variants of the language" (Wurm and Mühlhäusler 1985, Mühlhäusler). As great the numbers affected by this extension may be - and we cannot be certain of the numbers actually affected - it did not cover the entire population. Accounts such as those by Mead (1956, 371) for Manus Island and Orken (1954, 863) for the Rabaul-Kokopo area, both quoted by Mühlhäusler (1985, 58), confirm that a considerable amount of young men were unable to acquire Tok Pisin. He in turn gives the following reasons for their lack of proficiency:

- the disruption of the plantation economy and labour recruiting
- the destruction of most missions
- the deterioration of the country's infrastructure

Many of these factors, however, would change significantly during the Post-War period and beyond.

3.6 The Post-War Period

After World War II had ended, administration of the former Territory of New Guinea and of Papua once again passed to Australia in 1946, which merged the two territories into one. Supportive of the independence movement that soon arose among the indigenous population, the Australian Government relinquished control in 1975, freeing the way for the foundation of the new nation

of Papua New Guinea.

According to Mühlhäusler (1985, 59ff.), the post-war period had consequential effects on the social environment in Papua New Guinea. He names the following factors as ranking among the most significant, especially as they relate to the continued development of Tok Pisin:

- Breakdown of the prewar caste society: upward social mobility became accessible to the local population, including English schools
- Decline in the importance of German missions: new missionaries would now come from English-speaking countries
- Increased contact between the Highlands and the coastal areas
- Development of urban centres: jobs were to be found in towns, as was status
- Use of English as the medium for school instruction
- Development of new media, including written communication

Among these factors, the first would have been the most significant. To quote Mühlhäusler (1985, 60):

> This atmosphere in which upward social mobility was no longer the exclusive privilege of the white colonisers but also within the reach of some New Guineans, was also bound to break down some linguistic barriers. In prewar times, it was not unusual to hear the opinion that indigenes who spoke English should be punished for their cheek ; English was the language of the masters and Tok Pisin that of the boys - hence the name ' tok-boi '. Now English education in schools was promoted and schools teaching English were opened in many new areas.

English, the original lexifier language of Tok Pisin, was now once again available as input. As far as new media goes, the post-war period saw a whole array of Tok Pisin newspapers begun (and some of them abandoned as quickly). Periodicals such as the *Bougainville Nius, Toktok bilong Haus ov Assembli, Nius bilong Yumi, Nu Gini Toktok and Kundu* ran between 1962 and 1982 with various success and duration. The most important periodical to emerge is certainly *Wantok*, founded by Catholics in 1967 and published entirely in Tok Pisin. Romaine (1992a, 50) cites an editorial from August 1970, in which the staff clarifies their target audience as speakers of rural, colloquial Tok Pisin:

> Tok Pisin em i tok bilong yu ... inap nau i gat kain kain tok i save kamap long tok pisin. Tasol husat inap ritim? Em i hapkas tok pisin tasol. Man i hatwok

long ritim. Dispela niuspepa Wantok em bai i spik olsem wantok tru bilong yu, em i no tanim tok pisin. Nogat. Em i tok olsem yu yet yu tok.'[4]

The history of Papua New Guinea as a nation and Tok Pisin as one of its most important languages does, of course, not stop at this point. Since the pages above cover the periods of time which are of most interest to the analysis below, however, I will stop the historical account at this point and delve into the internal history of Tok Pisin and its linguistic development.

---

[4]Tok Pisin is your language. There are already a lot of words coming into Tok Pisin now. But who can read them? That's half-caste Tok Pisin. It takes a lot of effort to read. This newspaper Wantok will speak as a true *wantok* [friend] of yours. It won't transform Tok Pisin. Never. It speaks as you yourself speak.

4 Internal Development

Of the various models describing the linguistic development of Tok Pisin, there are two which have had the most impact and received the most discussion in the field: Mühlhäusler's model of developmental stages and Keesing's model placing emphasis on the Oceanic substrate. I will first detail the first three developmental stages of Mühlhäusler's model, given that they cover the time period most crucial to the analysis below. Afterwards, I will point out where it differs significantly from Keesing's model as well as if and how, exactly, the two might be reconciled.

4.1 Mühlhäusler's model

In the developmental history first put forth by Mühlhäusler (1974) and since then detailed in his many works on Tok Pisin (see, for instance, Mühlhäusler (1979), Wurm and Mühlhäusler (1985), Mühlhäusler (1982), Mühlhäusler (2003)) he posits five developmental stages for Tok Pisin: the "jargon phase", followed by what he terms phases of "incipient stabilisation", "nativisation", "creolisation" and finally, "depidginisation and decreolisation". The corresponding time periods can very, very roughly be given as before 1860, between 1860 and 1883, 1884 to 1914, 1914 to 1940 and afterwards, intersecting with the historical developments described above as follows:

- Pre-1860: Pacific Jargon English (incipient contact: whaling, trading)
- 1860 to 1883: Incipient stabilization phase (plantations)
- 1884 to 1914: Nativisation/Expansion (German New Guinea)
- 1914 to 1940: Creolisation (World War I and Inter-War Period)
- 1940 to present: Depidginisation and decreolization (World War II and Independence)

As indicated above, the first three are of the most interest to me at the present time. As anyone familiar with the emergence of contact languages is aware,

though, such dates are not the points at which abrupt changes alter the language in their then-present stage, but serve as rough guidelines for their overall development during what has to be seen as a constant, fluid process, a point which Mühlhäusler (1985, 36) already acknowledges: "[...] these classifications [...] are at too high a level of abstraction to be regarded as true mirrors of the development of Tok Pisin as a whole". "Their main weakness", he says, "lies in their being unable to cope with non-linear developments in the history of Tok Pisin [...].' One of these non-linear developments is caused by the fact that Tok Pisin or its earlier pidgin form did not develop at the same pace across the regions of Papua New Guinea and the Melanesian region as a whole, as Mühlhäusler illustrates in the following schema (adapted from Mühlhäusler 1985, 76):
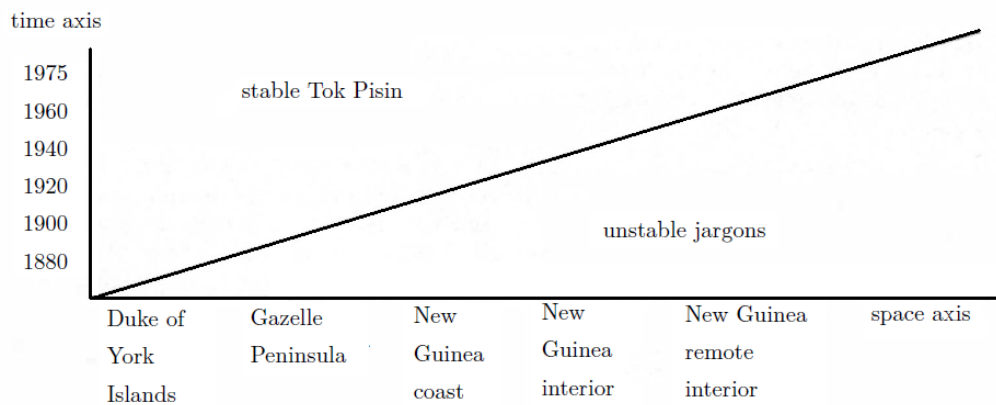
Figure 4.1: Time and space axis of Tok Pisin's development

Note that while the graph shows a linear development, said development is an ever increasing distribution of Tok Pisin across the regions of Papua New Guinea over time. However, the development and distribution took place on the Duke of York islands first, and in the remote interior of the mainland last. Hence, early developments in the structure and inventory of Tok Pisin would have taken place in the regions further left on the graph, while later developments could have taken place in more regions. Still, in the interest of looking at the development of the language as a whole, such abstractions will be necessary. Thus, let us take a look at how Tok Pisin, or its earlier forms, arose during the Jargon stage in Mühlhäusler's model.

4.1.1 Jargon Stage

Mühlhäusler distinguishes jargons in general "from proper pidgin languages by their excessive instability, extreme impoverishment in their expressive power and their high context-dependence" (Wurm and Mühlhäusler 1985, 80). He warns that "[d]ata on the jargon stage are scarce and what has been recorded tends to be biased in favour of an English character of the language", especially given that what little data are present stem mostly from the recollections of European voyagers. Consider this following excerpt from Thomas Jefferson Jacobs' account of his travels. Published in 1844, it is, according to Mühlhäusler (2003, 35), "the earliest example of Pidgin English spoken in the area of present-day Papua New Guinea".

> Presently the consultation ended, and the savages stood along the beach in battle array, while a noble-looking red warrior advanced to the water's edge, and, shaking his spear at us, cried out at the top of his voice, 'You make lie to kill us! You killed and eat Darco many moons ago! We know you, Potigo, very bad. Can't kill us on land! We kill you! You afraid of magic stone!' With this the savages uttered the war-yell and brandished their war implements, while the turn-turns were beaten with increased fury. Darco again hailed them: 'Me no speak lie! Me real Darco. Pongo good man; no eat me! Me hab been to America! Me come ashore alone, and show you! 'You speak lie plenty!' shouted the red warrior. 'Te-lum-by-by Darco not white like you!' 'Me not white!' shouted Darco, as he stripped off his duck shirt and trousers, and hung them on the jib-stay, upon which he seized, and stood up in a commanding attitude, exposing his bare body full to the view of the savages, while he stretched out his muscular arm and pointed at them with his hand, and shouted, 'You see me. I am Prince Darco!' (Jacobs 1844, 80)

Grammar and Syntax

Characteristics of this jargon included changes to the English input on all levels of grammar, according to Mühlhäusler (1985, 81ff.). Phonology exhibited extreme variation, with little effort made by Europeans to accomodate to the natives' pronunciation by avoiding difficult consonant clusters, voiced final stops or affricates, as indicated by documents from the times. Mühlhäusler (1985, 82) concludes that "as a result of the variation and insensitivity of jargon speakers for the sound patterns of their interlocutors, misunderstandings, mishearing and noncommunication must have been frequent." It is such miscommunication that may have formed the basis for some of the reanalysis of

English constructions in Tok Pisin, as we will see below.

In terms of morphology, the jargon was characterized by the somewhat expectable loss of inflectional and derivational affixes. Mühlhäusler (1985, 82f.) is careful in pointing out that this did not necessarily affect only the Pacific Islanders, but may also have been present in the Foreigner Talk of the Europeans (and therefore in the input the Islanders would have been subjected to). For instance, '"the data indicate that verb morphology was most consistently omitted by speakers of the lexifier language [...] [which] may have been a feature of the special register of foreigner talk used by the sailors to address 'natives'" (1985, 82f.). Similar arguments can be made for the loss or inconsistent use of the plural -*s* and the "considerable variation in the use of pronouns".

Syntax showed the same variation and inconsistency as phonology and morphology did. Mühlhäusler, in analysing what data we have from the era, concludes that "the overall impression [...] is that we are not dealing with a fixed language" (1985, 84), but with a jargon showing considerable variation in word order, the marking of word classes and the treatment of complex utterances; various solutions for translating items such as *very*; incomplete sentences, heavy borrowing from local languages and an absence of prepositions. He also posits an "iconicity of early Tok Pisin: the sequence of the elements of a sentence mirrors the sequence of events in the real world" (1985, 85). In a general evaluation of the evidence of syntax, he claims that "what little syntax there is tends to reflect universal strategies for discourse structuring rather than grammars of individual languages, or common denominators".

Lexicon

The lexicology of early Tok Pisin is, in Mühlhäusler's view, best "approached from two complementary points of view, first its reduction in size in comparison with its source languages and secondly the breaking down of lexical structures found in its source languages" (Wurm and Mühlhäusler 1985, 85). In terms of the first point, the number of lexical items in Pacific Piding English would have been very low. Concrete estimations range from 50 words (Speiser 1913, 50) to "about 300 words" Churchill (1911), but "are of limited relevance since none of them is based on an actual word count and since, furthermore, the size of the lexical inventory must have difered from locality to locality with changing requirements for verbal communication" (Wurm and Mühlhäusler 1985, 85). Topics covered by the vocabulary present are likely to have included the bare necessities of establishing trades, chiefly nouns, such as names for trade objects,

followed by other basic nouns, as Cassidy (1971) theorizes for pidgin languages in general:

> 'Natural materials (water, stone, fire, etc.); Physiography (mountain, river, sea, etc.); Time(yesterday, month, midday, etc.); Numbers (at least to ten, and some multiples); Weights and Measures (bundle, jarfull, arm's-length, etc.); Colors (black, white, and a few more of broad range); People, kinship (father, sister, son, child, chief, hunter, etc. - a large group); Body parts (head, hand, heart, eyes, teeth, skin - a large group); Weapons, utensils (spear, gun, stool, pot, hammer, bottle - a very large group); Clothing (shoes, shirt, dress, bracelet, etc.) ; Foods (general word, meat, oil, sugar, beverage, etc.); Animals, birds, plants, fish; Buildings (house, door, cart, bridge); Emotion, morality (fear, joy, lying, theft).' (Cassidy 1971, 214)

Verbs are likely to have included material to express basic concepts such as thought, communication, actions, desires, etc. Mühlhäusler (1985, 87) also notes that as was the case with the grammar of Pacific Pidgin English, its' vocabulary exhibited a great deal of variation, "depending on the inventiveness and experience of individual speakers, the composition of the groups using Jargon English and other factors".

The second factor in lexical change between input and the resulting jargon is given by Mühlhäusler as a "breaking down" of the lexicon of contributing languages, although he admits that "[t]here is no unambiguous statement about what is meant by 'breaking down' with regard to the lexicon" (1985, 87), suggesting "that those perceived distortions occurring when lexical items of English origin are used by the learners in a jargon context can be associated with a wholesale loss of lexical information, the loss of information about lexical relatedness leading to the disappearance of internal structure of the lexicon". Since pidgin is acquired at a very rapid pace, there is no chance to properly acquire the whole structure of the lexicon, both in terms of its individual 'entries' as well as their relation to each other, resulting in a "small and sometimes distorted subset of the lexical in formation contained in the lexical items of the lexifier language" (1985, 87). Mühlhäusler sees this distortion both on the phonetic level, in the loss of syllables, consonant clusters and phonological distinctions, and on the semantic level, with the "loss in referential potential" as "the most striking".

4.1.2 Stabilization stage

The stabilization stage is where Mühlhäusler's model of Tok Pisin development diverts most from Keesing's theory (see chapter 4.2) below). Mühlhäusler suggests that stabilization of its earlier pidgin form began "among the ethnically and linguistically diverse plantation workers on the Samoan plantations of the German Trade and Plantation Company, and subsequently on the plantations belonging to various German firms in the Bismarck Archipelago and the New Guinea mainland" (Wurm and Mühlhäusler 1985, 90), citing instutionalisation as a means of communication, continuity in transmission, withdrawal of English as a model language and emergence of standards of correctness as the driving factors behind stabilisation. Linguistically, he claims, stabilisation led both to the diminishment of lexical and grammatical variability and the emergence of structure in lexis and syntax. The following is the record of an early phonogram recording dated to 1904, "recorded in Monumbo on the New Guinea mainland in 1904 by Dr. R. Poech, a German doctor" and provided in Mühlhäusler (2003, 45).

> Belong place belong me me shoot him plenty kumul. Pass me come 'long place 'long white man, place he no strait. Plenty mountain. Pass ground he sitrait, me shoot him plenty. Me look him, he run away finish. Pass he top good, me shoot him. He karapaim long diway, me no look him. Pass me mark him long time, he run away, he go an other fellow diway.[1]

Note how, in contrast to the pidgin text above, it features, for instance, grammatical items such as the prepositional marker *belong* in *place belong me* (my area) and an early version of the transitive verbal marker *-im* in *shoot him*[2] and *karapaim*.

Grammar and Syntax

Phonologically, early stabilised varieties of Melanesian Pidgin still exhibited a lot of variation. Those variations which were detrimental to intelligibility

---

[1] Mühlhäusler's translation: If I were in my own area, I would be able to shoot many birds of paradise. Now I have come to the White man's area, this place is not flat, many mountains; I think one will not be able to find birds of paradise and shoot them. If the ground (?) is flat, I shoot many. But the ground is not flat. If the ground were good (?) I would shoot plenty. I hear them call out, but when I go towards them, I hear them (?) fly up high in the trees. If they stayed down low I would shoot many; but they do not stay down low, they stay up high in the trees. And they do not stay still in the trees; if I take aim at them, they fly away. If they stayed still in the trees I would shoot many.

[2] Note that while Poech gives 'shoot him', while Mühlhäusler (2003, 46) claims the pronunciation to be "something like ['ts uts im]".

would have been eliminated over time, but "the gap between this variety and that spoken by native speakers of English remained and, to some extent, exists even today" (Wurm and Mühlhäusler 1985, 91). As far as inflectional morphology goes, the stabilization phase saw the emergence of the transitive verbal marker *-im* and the adjectival marker *-pela*.[3] The most significant syntactic development during this stage was the decline of variation in word order in favour of a more rigid one, usually that of a subject-verb-object order. In parallel, and likely enabled by a more predictable word order, grammatical word classes start appearing. The predicate marker *i* also emerges during this phase, "its main function at this time appears to be to reinforce the syntactic information already expressed through invariable surface structure of elements" (Wurm and Mühlhäusler 1985, 94).

In addition, temporal deixis and other aspectual information is now encoded by sentential adverbs. Mühlhäusler (1985, 94) cites constructions such as *baimbai mi kambek, mi kambek pinis* and *mi kambek* as examples for the encoding of future, past and present (with a zero morpheme or *nau*, alternatively) as examples. In terms of other aspectual information, he quotes examples from Brenninkmeyer (1924, 23-24) as follows:

- iterative: *mi go planti taim* 'I go many times'
- intensive: *he tok planti* 'he talks a lot'
- wish: *i gut yumi go* 'let us go'
- adversative: *masi, yu mekim* 'why don't you do it?'
- frustrative: *yu go nating* 'you went in vain'

Variation in negation is diminished as well, with strategies such as double negation or the negation of individual constituents disappearing in favour of clausal negation. Complex and embedded sentences do not seem to be widely possible at this stage, only coordinative sentences with *nau* 'and/or' and *tasol* 'but' (Wurm and Mühlhäusler 1985, 95). An exception is subordination by means of *sapos* 'if'.

Stabilized Tok Pisin also included a stable pronoun system, which mostly remains unchanged in modern Tok Pisin. Table 4.1 below, adapted from (Wurm and Mühlhäusler 1985, 96), illustrates this fact:

mi	I, me
yu	you (sg.)
em	he, him, she, her, it
mipela	we (excl.)

---

[3]For details on their emergence, see chapters 6.4 and 6.3, respectively.

yumi	we (incl.)
yupela	you (pl.)
(em)ol	they

Table 4.1: Pronoun system in stabilized Melanesian Pidgin

Further syntactical complexification and elaboration would occur in the next stage, dubbed the expansion stage by Mühlhäusler.

Lexicon

The stabilization phase brought, along with a steady increase in the number of lexical items stemming mostly from borrowing, the following general lexical changes (Wurm and Mühlhäusler 1985, 97):

- the emergence of norms as to what constitutes a lexical item of the language;
- the crystallisation of preferred norms of lexical variants;
- the development of lexical field structure

All of these developments were influenced by the absence of English as the original lexifier language, so that "more often than is usual for pidgins, one encounters differences in phonological, syntactic and semantic information of Tok Pisin lexical items and their related English etymons" (Wurm and Mühlhäusler 1985, 98). The examples Mühlhäusler gives include phonological differences as well as morphological ones, such as plural noun forms being borrowed to refer to single entities (e.g. *anis* for ant (sg.)), English compounds being borrowed as simple bases (e.g. *bilinat* from *betel nut*) or the fusion of two or more lexical items into one lexeme (e.g. *nambis* (beach) from *on the beach*).

Differences in syntactical information between stabilised Melanesian Pidgin and English included a difference in the cases associated with with certain verbs as well as the position of adjectives. The position of the latter, i.e. whether they occurred postnominally or prenominally, depended on the adjective in question. According to Mühlhäusler (1985, 94) *bikpela, longpela* and *raunpela* would precede the noun, whereas *daun, kais* or *kela* would follow it.

Semantically, the lexemes of stabilized Melanesian Pidgin differed not only from English, but also from substrate languages such as Tolai. Tolai *virua* (victim, human flesh), for instance, becomes *birua* (enemy warrior) in Melanesian

Pidgin. Mühlhäusler (1985, 99) sees this development of the "meaning of lexical items independent of those found in the lexifier language as a further sign of Tok Pisin's status as an in dependent linguistic system".

By "the development of lexical field structure", Mühlhäusler understands a "way of organising the lexical material borrowed from a number of sources, thereby reconciling the frequently conflicting semantic information 'picked up' from these sources" (1985, 100). He gives the examples of the enumeration system (*wanpela, tupela, tripela* and so forth) as well as kinship terms. However, as he acknowledges, both these exemplary fields still show a measure of variation. Enumeration, for instance, was expressed by some speakers using a quinary system based partly on bodyparts (e.g. *wanpela han na tripela* 'one hand and three' for eight). Kinship terms, too, were only really stabilized in their central meaning, which did not necessarily match the meaning of the source language the lexeme was taken from, with *brata*, for instance, describing "a sibling of the same sex" (Wurm and Mühlhäusler 1985, 101).

In terms of lexical expansion, circumlocution played a central role. While many of the more complex circumlocutions were eventually replaced by shorter expressions, some, like *sit bilong binen* for honey or *rop bilong su* for shoelace remained in use (Wurm and Mühlhäusler 1985, 102).
Mühlhäusler's conclusion regarding the stabilization phase of Tok Pisin in his model is as follows:

> A stabilised pidgin, in the technical sense, is a pidgin which is governed by social rules and conventions in a limited domain of human discourse. Its primary function is that of a tool for exchanging information (referential function) rather than of expressing the full range of individual feelings and relationships between individuals and society as is the case in more complex languages. Because of its limited functional range a stabilised pidgin is reduced in its lexicon and its grammatical possibilities when compared to languages spoken natively or pidgins used for more complex purposes. (Wurm and Mühlhäusler 1985, 102)

4.1.3 Expansion stage

In Mühlhäusler's model, the expansion stage sets in during the time of the German administration. Several historic developments described above provided continunous stimuli for the expansion process, including the gradual pacification of the islands, the missionary efforts and the Second World War and its aftereffects.

The following text example is quoted from Hall (1943b, 40), and is also included in Mühlhäusler et al.'s collection (2003, 59)[4]. Hall titled it "a punitive expedition".

> Nau bipo longtaim mi polisboi, mi stap Ambunti. Orait, mi stap gutpela. Mi no gat trabol. Olo kanaka bilong bus olsem i no stap. Oltaim oltaim ol i pait. Orait. Baimbai nambawan i harim kiap. Nambawan kiap i harim pinis, i salim pas long Stesin. Nambatu kiap kisim pas finis, bihain em i tokim mipela. Em i tok, mipela olgeto go long bus bilong lainim olgeta kanaka. Orait. Bihain, tupela de mipela go nau. Mipela go go go long kanu, tudak. Tudak, orait, mipela slip nau. Slip finis, long moningtaim mipela kirap.[5]

Mühlhäusler (1985, 106f.) sums up the expansion stage through four key developments. First, there was an expansion of the usage contexts Tok Pisin was employed in, which now included "integrative and expressive purposes". As speakers increasingly used it to express feelings and desires, it became part of their identity, even though it was not their native language. Secondly, the usage in official and business domains increased as well, including such contexts "as religion, economy, agriculture, education, aviation, modern warfare, and parliamentary transactions". Thirdly, the media began to use Tok Pisin, including the radio, pamphlets, newspapers, books and, in later times, films and theatre plays as well. Finally, Mühlhäusler notes, "Tok Pisin today no longer is supplementary to the traditional vernaculars but is beginning to take over their functions, thus leading to the functional and structural decline of vernaculars in some areas".

All of these societal changes included, of course, changes to the structure of Tok Pisin itself, which are detailed in the sections below.

Phonology

In terms of phonology, there were different developments regarding consonants as opposed to vowels. For the latter, two distinct systems emerged in the

---

[4]For purposes of easier reading, the version given here is the orthographic transcription as it appears in Mühlhäusler (2003, 59) rather than the phonemic transcription Hall uses.

[5]Now long ago I was a police-boy (Native policeman), and I was at Ambunti. I was always good and did not make trouble; but (all) the backwoods Natives were not always thus. They were (all) continually fighting. Finally the government official in charge got word of it. When he had heard about it, he sent a letter to the post. When the subordinate official received the letter, he spokes to us. He told us all that we were to go to the backwoods in order to each (or line up in allegiance) (all) the Natives. Then we went for two days. We kept going until night. When night came, we slept; and when we had slept, we got up in the morning.

geographical varieties of Tok Pisin: those varieties spoken mostly in the interior areas of the mainland exhibited a five-vowel system ([i], [e], [a], [o], [u]), whereas some coastal varieties exhibited first seven vowels ([i], [e], [ɛ], [a], [ɔ], [o], [u]) before settling on a ten-vowel system including [a], [aː], [e], [ɛ], [i], [ɪ], [o], [ɔ], [u] and [ʋ].

As regards consonants, however "irrespective of a Tok Pisin user's native language, a number of distinctions which were not encountered in earlier Tok pisin are now widely made [...] includ[ing] a distinction between [s] and [t], [p] and [f], and [l] and [r] in that order" (Wurm and Mühlhäusler 1985, 108), so that "the consonant system of expanded Tok Pisin can be summarised as follows":

		Labial	Labiodental	Dental	Palatal	Velar	Laryngeal
Plosive	Unvoiced	p		t		k	
Plosive	Voiced	b		d		g	
Nasal	Voiced	m		n		ŋ	
Fricative	Unvoiced		f				
Fricative	Voiced		v				
Sibilant	Unvoiced			s		(š)	
Sibilant	Unvoiced					(č)	
Affricate	Unvoiced					(ǧ)	
Affricate	Unvoiced						
Trill	Unvoiced			r			
Lateral	Voiced			l			
Aspiration	Unvoiced						(h)

Table 4.2: Consonant system of expanded Tok Pisin

Phonological rules also expanded, allowing for some consonant clusters that violate the preference of CVCV structures for earlier stages (Wurm and Mühlhäusler 1985, 109). As Mühlhäusler puts it, "the relaxation of phonological restrictions appears to be directly responsible for the increasing number of phonological rules, in particular rules which allow speakers to de-emphasise marginal semantic information and rules for allegro pronunciation".

Grammar

In terms of inflectional morphology, expanded Tok Pisin still shows no innovations, although Mühlhäusler (1985, 111) speculates on possible future

affixes, claiming that "[i]t is conceivable that em i may in future become phonologically reduced to yield a prefix of the type found in a hypothetical *man mi kam where mi- fulfils the role of the present-day predicate marker i".

The majority of developments between stabilised and expanded Tok Pisin, however, took place within the area of syntax, as we will see below as well. Mühlhäusler (1985, 112) summarises these developments as follows:

- the emergence of compulsory grammatical categories;
- the development of embedding and overt signalling of embedded constructions;
- changes in word order for grammatical and stylistic reasons;
- the development of discourse structuring devices'

The examples Mühlhäusler gives for a) are sentential adverbs enabling expression of tense, aspect and number, such as *baimbai*, which eventually becomes an "obligatory preverbal aspect marker *bai*". Another is the nominal plural marker *ol.* The latter of these examples are dealt with in detail in the analysis below. As for the motivation of this development, Mühlhäusler claims that "the introduction of obligatory and redundant categories must be seen as decreasing the context sensitivity of the language and at the same time making up for the loss of perceptive power caused by the development of phonological rules".

In terms of the development of embedding and associated markers, expanded Tok Pisin includes several means for marking embedded sentences with relativizers and complementizers. Again, the individual complementizing strategies will be dealt with in detail in chapter 6.11. For now, it will suffice to say that Mühlhäusler sees their origin in sentence adverbials (*olsem* and *baimbai*), prepositions, (*long* and *bilong*) and verbal concatenation (*se*).

In his view, there are two possible reasons for developments in word order of a pidgin, either to "delineate the scope of grammatical particles such as negators or aspect markers [or] to allow the movement of semantically prominent material into prominent syntactic positions" (Wurm and Mühlhäusler 1985, 118). Examples for the former include markers moving closer to the syntactical constituent they modiy, such as *bai* moving closer to the verb and *ol* moving closer to the noun. For the latter, Mühlhäusler gives the following example as a means of using word order to emphasize syntactical constituents:

(1) planti snek mi   lukim
     lots   of   snakes I   saw

(2) mi lukim planti snek
     I   saw   lots   of   snakes

Variant (1) would be marked in fronting the object, whereas (2) would have been the expected word order. Note that both variants would have been possible in earlier, stabilised Tok Pisin as free variants.

Lexicon

Mühlhäusler (1985, 119) posits two main developments in word formation, which happened at different points in the expansion stage:

- Early expansion stage: increase of word-level compounds. As a rule the surface structure of derived lexical items is relatively close to their putative deep structure (e.g. guttaim (from gutpela taim *good time*) *peace* or lukbuk (from lukim buk *to look at a book*) *to read*).
- Late expansion stage: strong tendency to derive word-level rather than phrase-level lexical items, increasing discrepancy between lexical surface structures and related deep structures, lexical programs becoming increasingly productive.

In general, the expansion stage sees a vast increase in lexical productivity. As Mühlhäusler (1985, 120) remarks, stabilised Tok Pisin still had few to no resources to create new words internally, relying mostly on borrowing instead. This changes drastically during the expansion stage and later stages, with modern Tok Pisin having "about 25 programs or patterns involving multifunctionality [...], about 25 programs of compounding, and about a dozen programs which use reduplication to create new lexical items".

For the growing tendency of word-level derivation, Mühlhäusler traces the development from the construction *man bilong V$_{ing}$* to *Nman*, with earlier forms such as *man bilong pait* (fighter) or *man bilong pret* (fearful person) replaced by shorter, word-level forms *paitman* and *pretman*.

Stylistic expansion

A final note on the expansion stage regards stylistic devices becoming available to speakers of expanded Tok Pisin. Mühlhäusler (1985, 127) sees "one of the principal effects of grammatical expansion of Tok Pisin [in being] that it provided structural and lexical alternatives which could be employed for stylistic

purposes". He lists strategies such as 'backsliding' (consciously using less developed varieties) and 'sidesliding' (using variants associated with a different part of the restructuring continuum). Another option open to speakers of expanded Tok Pisin was to choose between the various competing constructions, such as using relativizer *we* and *ia*-bracketing, which served the same function in different stylistic contexts. Similar effects could be achieved by using the various forms of *baimbai, bai* and *ba.*

This concludes the summary of the three stages of Mühlhäusler's model as relevant to the analysis at hand. I will now move on to cover the key differences between Mühlhäusler's developmental stages and the Oceanic substrate model as put forward by Keesing.

4.2 Keesing's model

The main difference between Keesing's model and Mühlhäusler's model as explained above is that Keesing posits a much closer link between the pidgin of Queensland and areas of the southwestern pacific, as well as a corresponding much earlier stabilization of what would eventually become Tok Pisin, as early as 1850s. He discounts the importance of the plantations, locating the stabilization "in shipboard settings and trading enclaves, not on plantations" and argues that this precursor was much more widespread and uniform than Mühlhäusler and colleagues argue. He sums up his position in his own words as thus:

> My reading of the historical and linguistic evidence is that Mühlhäusler (1977, 1985a, 1986) and before him Salisbury (1967) have radically overestimated the separateness of the New Guinea Pidgin lineage from the pidgin of Queensland and the recruiting areas of the southwestern Pacific. I read the evidence as indicating that the pidgin spoken both on the German plantations in Samoa and in some parts of the Bismarck Archipelago as of the first half of the 1880s was essentially the same as that spoken in Queensland, the New Hebrides, and the Solomons: these areas were part of a single, dispersed speech community. Only in the latter 1880s—and only after Pacific pidgin had acquired a striking degree of expansion and stability—did New Guinea Pidgin undergo the special developments, including relexification from Bismarcks languages, that distinguish Tok Pisin from other daughter dialects of Pacific pidgin. (Keesing 1988, 3f.)

Furthermore, he believes that these "special developments" took place during earlier stages than most other linguists have claimed, and took place not on New Guinea, but hundreds of miles away. Rather, he locates the nexus of the origin of a "distinctive Pacific pidgin [...] in the central Pacific, in a zone where pidgin has not been used for many decades". This origin, he claims, is to be found not on the plantations of Queensland or elsewhere, but on ships and in trading enclaves. The plantation setting, according to Keesing, was not relevant to the stabilization, but to the relexification of Tok Pisin later on.

By necessity, he continues, "this distinctive Pacific pidgin [...] was primarily a medium of communication among ships crews [sic!] of mixed-island origin" and was "used across a vast expanse of ocean on which the islands are tiny, scattered dots". While admitting that it was not yet fully stable, he claims that it was relatively uniform by a few decades after 1845.

The split that divided this uniform pidgin into distinct varieties did not occur until the 1880s, as per Keesing:

> Local dialects of this pidgin began to emerge in the 1870s, when plantations replaced ships and trading enclaves as the main venues for the expansion of the developing regional pidgin. French, German, and indigenous languages of Queensland, Samoa, Fiji, and the Bismarck Archipelago added locally distinctive lexical content to a common lexicon derived from nautical pidgin (and such specific regional forms as the pidgin of the China coast) and directly from English. But I believe these local lexical innovations were, until the end of the 1880s, only minor elaborations and modifications of a common regional pidgin, variations quickly learned by the substantial numbers of Islanders who moved from one plantation area to another as recruiters, foremen, interpreters, cultural brokers, and multiple-term laborers. (Keesing 1988, 4)

In regard to substrate influence, Keesing argues that "Melanesian Pidgin is in a curious position in relation to the lively arguments [...] stirred to a boil by Bickerton's recent work", acknowledging the "obvious parallels between Melanesian Pidgin and Oceanic languages", such as the inclusive/exclusive distinction for first person nonsingular pronouns. He disputes their explanative power for structural development, however, noting that "these arguments seem to be possible whatever indigenous language is chosen for comparison" and that similar substrate influences occur in contact languages far removed from the Melanesian area. At the same time that its origins are claimed as proving substrate hypotheses, he goes on to criticize, Tok Pisin "has also provided material for universalist arguments, particularly as it has undergone grammatical and lexical elaboration, standardization, and finally nativization", citing Gillian

Sankoff's work on *baibai* becoming a preverbal aspect marker, *i* becoming a predicate marker and the development of relative clauses. "All this is not to deny the possibility of substrate influences from New Guinea langauges, both Papuan and Austronesian", he assures, "but we must recognize that this is not the place to look for the sources of grammatical patterns common to Tok Pisin and the Vanuatu and Solomons dialects". That place, he claims, is the central Pacific area.

More specifically, he locates the settings for the first expansion stage of what he calls a "developing Pacific lingua franca [...] to have been a series of interlinked island groups, principally Pohnpei (Ponape) and Kosrae (Kusaie) in the Carolines, the Gilbert Islands, and Rotuma, which were favored venues for whalers, traders, beachcombers, and deserters", islands central to the recruitment of sailors for both whaling and trading vessels. He specifically disputes Mühlhäusler in the quality and quantity of contact between Europeans and such sailors:

> Mühlhausler, in a sketch of the history of English-based pidgin(s) in the Pacific (1985a: 38), comments that prior to 1860, "Islanders serving on board European vessels, and Europeans deserting to the islands and living among the natives are the exception rather than the rule." My reading of the evidence suggests that Mühlhäusler is wrong. The total population of Islanders serving on European vessels was quite substantial enough to constitute a potent force of linguistic (and cultural) innovation and transmission, particularly insofar as they communicated with one another as well as with Europeans, and insofar as their wanderings to distant places gave them a prestige and sophistication, and a role as linguistic and cultural brokers, far beyond their absolute numbers. (Keesing 1988, 22)

It is the role as "linguistic and cultural brokers" the sailors served, in Keesing's view, that allowed a Pacific pidgin to spread far beyond the ships themselves and establish itself as a uniform regional lingua franca.

As far as the latter complexity and elaboration of that pidgin into what would become Tok Pisin and the other Melanesian dialects, Keesing discounts both Bickerton's (1984) later suggestion that it was the plantation period and the transmission over several generations which provided the necessary setting for such expansion as well as a nativization and repidginization scenario following a phase of radical expansion. Here, he admits that "[t]hose Islanders who acquired childhood fluency in this language of shipboard life and for whom it may have been the main language [...] would have been relatively few in number, and scattered", acknowledging that they would have lacked the social

and political power to exert much influence on the emerging language. He maintains, however, that they "played a crucial part in its dissemination [and] served as models for adult language learners" both indigenous and of European descent.

Just like Keesing assumes earlier dissemination and stabilization, he also assumes that the elaboration of Melanesian Pidgin happened earlier, and is "a late nineteenth century phenomenon" rather than one occurring during the twentieth. In his view, Tok Pisin did not undergo steady development towards an extended pidgin, but instead suffered setbacks in this regard. It is the "special case of the transplantation of an extensively elaborated pidgin to alien linguistic soil" that caused it to undergo "considerable withering of its syntactic resources", he argues. What Mühlhäusler attributes to its first elaboration, he terms "progressive reconstitution and of an already elaborated pidgin", in addition to "partial relexification from a new source substrate language", Tolai, and "a new superstrate language" (German). These new influences replaced earlier "relatively clear substrate models and motivations [...] when we look at the setting to which the New Guinea dialect was transplanted", causing "many patterns pervasive in EO languages [to be] absent".

To summarize, then, Keesing believes that shortly after 1845, a uniform Pacific Pidgin stabilized, which remained until the 1870s. Next, this dialect split into the three Melanesian dialects of Bislama, Tok Pisin and Solomon Islands Pijin. He locates the center of this stabilization not on the plantations but on the ships, yet argues that the pidgin still remained uniform through the Central Pacific. As maybe his most important structural point, he claims that the grammatical features present in all three Melanesian dialects derive from Oceanic substrates.

4.3 A combined model?

There are three main potential issues I see with Keesing's theory. First, I doubt that the situation and numbers of Pidgin speakers on board of the ships spread across the Central Pacific region was conducive to stabilization of a pidgin. Secondly, there is considerable evidence that if substratum influence was involved in the emergence of Tok Pisin, it was not Oceanic substrates that shaped the language, but Austronesian substrates. Third, Keesing's observation that the three Melanesian dialects share some distinct features does not automatically indicate that these features stem from a common origin. I will go into all three of these points in detail below.

First, consider the numbers given above in chapter 3.2.1 by Bennett on the number of whaling contacts on the Solomons. They total less than 130 ships over a period of 88 years. Take in addition to that the fact that whaling ships were well known to have frequent crew rotation, and you see why Tryon and Charpentier (2004, 112) note that only a few Islanders would have worked aboard whaling ships. It would have to be somewhat later, probably during the trading period of the mid- to late 1850s, that numbers would have reached sufficient levels for a stable pidgin to emerge among sailors and labourers on the trading posts. This, then, would align with Keesing's theory that a uniform Pacific Pidgin emerged shortly after 1845, with exploitation of the sandalwood beginning around 1841. However, the main location of contact at this time was no longer limited to the ships themselves.

In addition, consider what the studies above on the relationship between linguistic complexity and the size of the speech community has shown. Larger communities lead to more quickly established and more consistent structure, whereas smaller communities are more fertile ground for high diversity and complex structures. This would indicate that the situation on the ships, with only small numbers of speakers per ship, we should see more variation, a lack of norm-enforcing structures and parallel innovation - all three the opposite of stabilization as suggested by Keesing.

Secondly, most of the structures linked to substrate influence in Tok Pisin have been linked not to Oceanic substrates, but to substrates from the Austronesian area. Romaine (1992a, 41) writes:

> Baker (1990) lists more than 60 features of Melanesian Pidgin attested far earlier in New South Wales Pidgin English or Queensland Pidgin English, which was initially an extension of the former as Europeans pushed north, than in Melanesian Pidgin English. Among the features cited by Baker which occur on average about 35 years earlier in New South Wales or Queensland Pidgin English are both lexical and grammatical items such as *gammon* 'to deceive, trick', *walkabout* 'to run away', *belong* (as a genitive), *by and by* (as a sequence marker [...]), *been* (as perfective marker), *fellow* (in seven syntactic slots/functions), and *him* (as a transitive suffix). This evidence is damaging to Keesing's claims about chronology as well as the alleged Oceanic source of these items in Melanesian Pidgin English. If substrate is crucially involved in the early formative stages as Keesing claims, then the origin of these features must be in the Aboriginal languages of Australia rather than Eastern Oceanic languages.

We will see below that this is also true for the constructions surveyed in this work, casting further doubt on an Oceanic substrate origin for Tok Pisin.

Thirdly, just because Tok Pisin, Bislama and Solomon Islands Pijin show similar or identical structures in certain regard, this is by no means conclusive proof for a common origin of these structures in an earlier, stabilized, uniform Pacific pidgin. In particular, Keesing claims that the following syntactic and lexical patterns are common to "all three daughter dialects of late 1880s south-western Pacific pidgin" (Keesing 1988, 112f.):

1. the pronominal system

2. the use of pronouns to embed relative clauses

3. the systematic use of the transitive suffix *-im*

4. the grammaticalization of TAM markers

5. the use of both periphrastic causatives using *mek-* and grammaticalized causatives using the transitive suffix *-im*

6. a number of complementizing and relativizing constructions including *see, wea* and *bulong*

7. the use of *-pela* and *-fela*

8. a set of clause-initial phrasal interrogatives

9. a distinct system of marking possession

10. *ia*-bracketing

As remarked above, some of these constructions could have arisen due to Austronesian, not Oceanic substrate influence. However, I would argue that some of these construction could also be, as Mühlhäusler claims as well, determined by universal tendencies of grammatical development. I will show below that the grammaticalization of TMA markers, the origin of complementizing and relativizing constructions as well as *ia* bracketing require no common origin, but can be motivated by internal processes within Tok Pisin that could have run parallel in Tok Pisin. Furthermore, these are more complex syntactic phenomena that would not have been largely present in the simple pidgin likely to be spoken as a nautical or trade pidgin.

Structures such as the transitive suffix *-im* and the use of *-pela* and *-fela* may be, as Keesing claims, derivatives of such an early pidgin, as will also be discussed below. The question to be answered here, then, is when to date stabilization of Tok Pisin, and upon what criteria.

The amount of lexical items available to speakers of the pidgin, for instance, is a necessary, but surely not a sufficient factor to determine the stabilization of the emerging language. Even if we were able to conclusively determine the exact number of lexemes in the average speaker's vocabulary, how would we determine which number is sufficient? As Romaine (1992a, 40) points out, Mühlhäusler (1979), for instance, claims that Tok Pisin was still lexically impoverished before the first World War, whereas Crowley (1990) and Clark (1987) argue that it had already expanded its vocabulary to around 1,000 words by the 1880s. Would this indicate that if, on April 27th, 1901, the average speaker had had a vocabulary of 894 items, it would be an impoverished lexicon? Or a stabilized one? Surely it can not be the quantity of the vocabulary that determines a pidgin's status as stabilized, especially not as the sole factor.

It may even be possible that both Keesing's and Mühlhäusler's models of stabilization have merit. Of course, they still have different implications as to the importance of different substrates. Romaine (1992a, 42) remarks that "Keesing is probably right that there was considerable stabilization at an earlier stage than most scholars have thought possible, but I am dubious about the extent to which an Oceanic substratum determined the structure of the Pacific pidgins, and the degree to which ships provided a stable base for crystallization of the pidgin".

I agree with this assessment. It seems likely to me that certain structures, such as the transitive marker, the predicate marker and the adjectival marker stabilized early. However, more complex syntactical structures such as complementizers would have neither been needed at that stage, nor would they have been introduced through Oceanic substrates. We will return to the question of when Tok Pisin actually stabilized - and whether it is at all beneficial to assume a period of stabilization - after reviewing the evidence below.

5 Methods and material

5.1 Z'graggen subcorpus

The empirical analyses below are based on the Z'graggen corpus. This corpus contains 4,645 text files in Tok Pisin as recorded by Fr. John Z'graggen from the early 1970s to the 1980s. John Z'graggen was a missionary and linguist - his dissertation was on the languages in the Madang District in New Guinea - so we can assume that they were gathered with linguistic analysis in mind. The original tapes the texts are based on are held in the Museum der Kulturen in Basel, Switzerland. The transcriptions themselves were graciously made available for this research through the PARADISEC archive (Z'graggen 2011). Overall, the corpus has, in its cleaned form (see below), 8,489,942 tokens and 35,419 types, which results in a type/token ratio of 0.00417. The texts are mostly narrative in nature, with the recorded speakers telling a story and the recorder sometimes posing questions to keep them talking. An exemplary text is reproduced below, along with my free translation:

> Bipo wanpela man nem bilong Okdum. Orait dispela ples, dispela graun i nogat. Graun em i ston tasol.
> Orait dispela man em i stap long hul bilong ston tasol, i stap na em i kaikai dispela samting ol i kolim Bulumu. Em i kaikai Bulumu na em tromoi pikinini bilong em antap long ston. Orait ol binatang ol i kolim anis bilong graun. Ol i stap aninit long dispela ston na kisim graun aninit long dispela stona nau putim i kam antap. Putim kam antap na i pulap antap long ston, na wokim maunten i go i bikpela maunten. Orait Okdum i tok nau bikpela ren i kamdaun na i brukim dispela graun i go. Na graun ia i go antap long ston na karamapim ol ston. Graun i go antap long ston na ston i stap aninit. Bipo graun i stap aninit na ston i stap antap, orait nau ston i go aninit na graun i kam antap. Em dispela stori i olsem mi harim.

> Long ago, there was a man called Okdum. Then, this place, this ground was not there. The ground was pure stone.

Allright, this man was living in a hole of pure stone, and he ate these things they call Bulumu. He ate Bulumu and he threw their seeds onto the stone. Then, insects called ants on the ground. They were underneath this stone and took the ground underneath this stone and put it on top. They put it on top and filled the top of the stone, and made a mountain that became a big mountain. Then, Okdum said now great rain comes down and breaks this ground in two. And the ground went on top of the stone and covered the entire stone. The ground went on top of the stone and the stone was underneath. Before, the ground was underneath and the stone was on top, and now the stone went underneath and the ground came on top. This story is the same I heard.

This text is typical for the data in the corpus. The speaker tells a story they themselves have heard before, relating it to the missionary. Many of the texts are the speakers telling stories not about themselves, but about mythical beings such as Manub and Kulbob, their ancestors, or times before men came to the islands, or the times before the white men came.

The texts of the recordings themselves are of rather differing length, as figure 5.1 shows. In spite of the discrepancy in word length, I made the decision to include all texts in the corpus so as to have the maximum available data to sample from. Individual samples were checked so as not to have an undue number of sentences taken from the same file.

5.1 shows where the texts fall on word length, with the horizontal axis being the identifier of each individual text and the vertical axis showing its word length. As is evident, there are a few texts up to around 15,000 words in length, while the majority of the texts are between 1 and 2,500 words long. In other words, the texts in the corpus show a great variety in length, which indicates that there would be some speakers who are overrepresented in the data.

5.2 shows the type/toke ratio of the individual recordings. It ranges from 0.03474745 to 0.7000000, with the vast majority or recordings ranging between 0.1 and 0.3, indicating a low treshold of lexical variety in the data.

5.1.1 Deduplication

PARADISEC supplied the transcriptions in the format of 4,645 .txt files. Unfortunately, some of these files contained duplicates of one another, requiring a process of deduplication before the corpus could be used in the intended
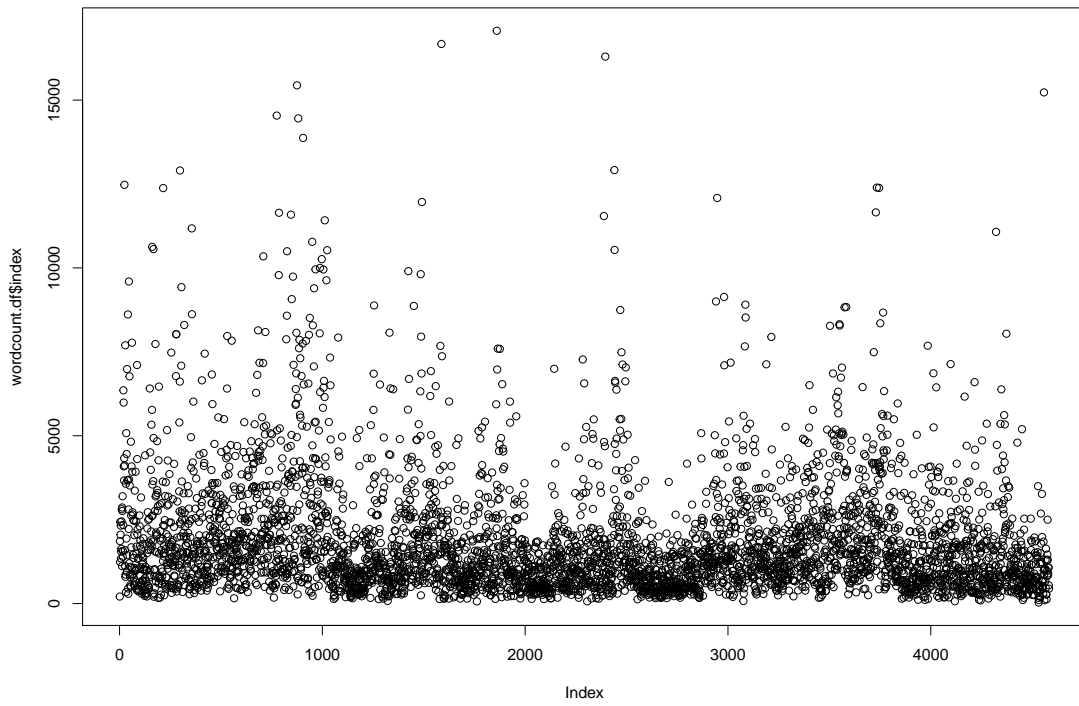
Figure 5.1: Word length of the text files
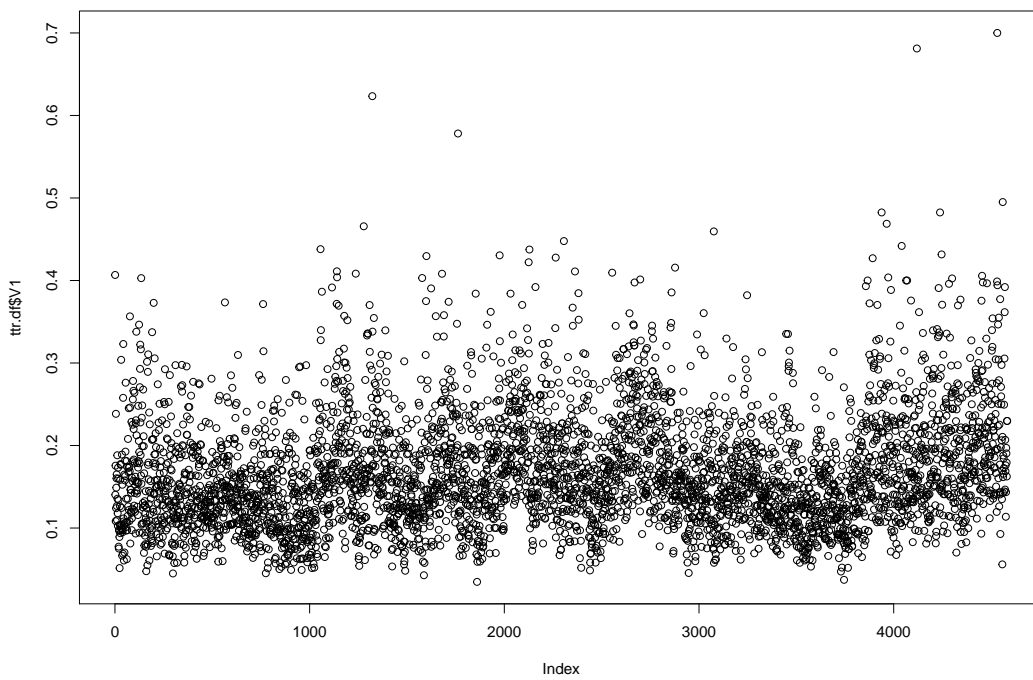
Figure 5.2: QQ plot of the type/token ratio of the text files

manner. Unfortunately, it was not a simple matter of some text files being duplicates of another, but some text files containing passages taken from another, while the rest of the text was unique. The following table shows an excerpt of the approach taken to deduplication [1]:

Duplicate search used	File 1	File 2	File 3	Match?	File kept	Files deleted
Orait nau yu bigin long as tru eh?	00001a	00174	04590_2	Yes	00174	00001a, 04590_2
Bipo yu no stori long mi, eh	00001b	00175	04590_3	Yes	00175	00001b, 04590_3
Orait stori long bipo long ol tumbuna papa	00001c	00173	04590_4	Yes	00176	00001c, 04590_4

Table 5.1: Excerpt from deduplication table

Table 5.1 above illustrates the process of deduplication. The first column indicates the first line of duplicate text as taken from DupliFind, an automation utility searching for duplicate lines in text files. All 5.058 text files produced by the first deduplication step were loaded into AntConc, and the duplicate line was searched across all of them. The numbers of all files that contained the lines were noted (see columns 2-4). The file size and the last lines of the text files were compared, and if they matched (column 5), the texts were checked manually for complete identity. In case complete identity could be established, the file with the highest text ID was kept [2] (column 6) and the ones with lower text IDs were deleted (column 7). All in all, the deduplication resulted in a lowering of the total token number from 9,403,011 to 8,489,942. In only four cases did the end of the text file and the file size not match. 01952 contained a shortened version of 02245, and was subsequently deleted. 00529 contained both the text of 02563 and 02581 (which were identical) in addition to other paragraphs. 02563 and 02581 were deleted.

Another six text files[3] were found to be empty, containing no text. In four other cases, parts of the text of one file were repeated in another file, while the rest of the file contained original text. The duplicate text was removed in the file with the higher text ID[4]. Furthermore, texts 00218 and 4178 contained paragraphs that were repeated in the text. These were also deleted.

---

[1] The deduplication table in its entirety is included with the supplementary data package for this publication.

[2] An exception is 04590, see below.

[3] These were 02595, 02809, 02810, 03264, 03265 and 02502.

[4] Text from 00432 was removed in 00833, text from 00685 was deleted in 00686, text from 04488 was removed in 00556, and text from 555 was removed in 742

```
# read files
setwd("~wd/corpus/clean")
input_files = list.files(pattern = '*.txt')
input_file_number <- seq(input_files)
library(tokenizers)
library(readr)

# extract sentences and create a new text file for each existing text file where one sentence is one line

for (fileName in input_files)
{
  setwd("~wd/corpus/clean")
  textin = read_file((fileName))
  sentences = tokenize_sentences(textin, lowercase = FALSE, strip_punctuation = FALSE, simplify = FALSE)
  bihain_long.ocs <- unlist(lapply(sentences, function(ch) grep("\\bbihain long\\b", ch, value = TRUE)))
  setwd("~wd/corpus/bihain long occurrences")
  write(bihain_long.ocs, file = "bihain_long_ocs_master.txt", append = TRUE)
}
```

Figure 5.3: Sentence extraction code in R

In addition, 00015 was deleted because it contained only English text. In general, any English text, usually metainformation such as titles announcing the content of the next section of the transcription, was removed.

5.1.2 Text processing and sampling

In a second step, the individual texts had to be processed into a format that could be read into Exmaralda (Partitur) for annotation. As described above, annotating the entire corpus was not possible due to its size and the available time. Instead, a number of samples was drawn for each of the phenomena under scrutiny, and, where necessary, these samples were annotated. In the interest of full transparency of he sampling process, I have included the R code below. It illustrates the first step in this process, which was extracting all sentences that included the form in question (in this case, the prepositional construction *bihain long* from all the texts in the corpus. The sentences with this construction were then written into a new text file (the master occurrence file). In the second step, the master occurrence file provided the basis for the sampling of a preset number of sentences that contained the construction in question. The code below illustrates how the master occurrence file was read in and remaining meta information not previously deleted was removed. The input was tokenized by sentence using the *tokenizers* package. As a final step before annotation, a random sample of the sentences was taken using R's sample function, and the result was written into the master sample text file.

```
setwd("~wd/corpus/bihain long occurrences")
library(readr)
library(tokenizers)

# read master occurrence file and remove lines with meta information

bihain_long.occurrences.master = read_file("bihain_long_ocs_master.txt")
bihain_long.occurrences.master <- gsub("\r", "", bihain_long.occurrences.master)
bihain_long.occurrences.master <- gsub("\n", "", bihain_long.occurrences.master)

# tokenize sentences

bihain_long.occurrences.master <- tokenize_sentences(bihain_long.occurrences.master, lowercase = FALSE,
    strip_punctuation = FALSE, simplify = FALSE)
bihain_long.occurrences.master <- unlist(bihain_long.occurrences.master, recursive = TRUE, use.names = TRUE)

# extract random sample of 250
master.sample.bihain_long <- sample(bihain_long.occurrences.master, 250, replace = FALSE, prob = NULL)
master.sample.bihain_long
write(master.sample.bihain_long, file = "master_sample_bihain_long.txt", append = TRUE)
```
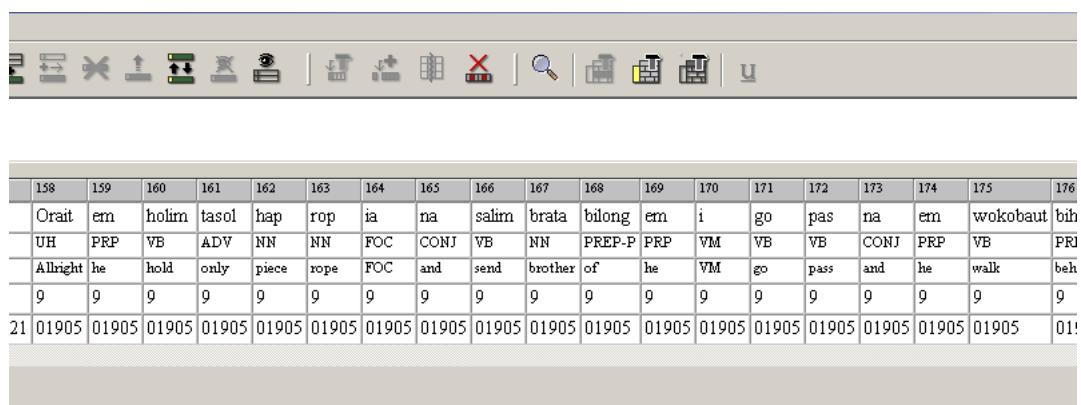
Figure 5.4: Sampling code in R

5.1.3 Annotation

The following screenshot gives an impression of the annotation process. The sample has been tokenized by Tok Pisin (from hereon abbreviated as TPI) words, with the uppermost row showing the token. From there, top to bottom, the rows show the syntactic annotation, the English translation, the sentence ID of the sample, and the text ID the sample was taken from.

	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176
	Orait	em	holim	tasol	hap	rop	ia	na	salim	brata	bilong	em	i	go	pas	na	em	wokobaut	bih
	UH	PRP	VB	ADV	NN	NN	FOC	CONJ	VB	NN	PREP-P	PRP	VM	VB	VB	CONJ	PRP	VB	PRI
	Allright	he	hold	only	piece	rope	FOC	and	send	brother	of	he	VM	go	pass	and	he	walk	beh
	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
21	01905	01905	01905	01905	01905	01905	01905	01905	01905	01905	01905	01905	01905	01905	01905	01905	01905	01905	01!

Figure 5.5: Exmaralda Partitur screenshot

Annotation tags used were as follows[5]

NN	noun
NNP	noun; proper
VB	verb
SV	serial verb
PREP-P	preposition; possessive

---

[5]Note that as the annotation was a step prior to analysis and writing, the tags in the annotated files may not reflect the glosses in the analysis chapter below.

PREP-C	preposition; comitative
PREP-I	preposition; instrumental
PM	predicate marker
FUM	future marker
PSM	past marker
ADJ	adjective
ADV	adverb
ADV-T	adverb; temporal
UH	interjection
FOC	focalizer
PRP	pronoun; personal
CONJ	conjunction
DET	determiner

Table 5.2: Annotation tags used

Annotation was done largely by hand. After the initial annotation period, the following items were judged to be common enough in a certain meaning or function that they were afterwards annotated automatically at the beginning of each sample annotation:[6]

Type	Item	POS	Translation
Conjunction	o	CONJ	or
Conjunction	na/Na[7]	CONJ	and
Noun	pikinini	NN	child
Noun	meri	NN	woman
Noun	mama	NN	mother
Noun	papa	NN	father
Noun	brata	NN	brother
Noun	hul	NN	hole
Noun	diwai	NN	tree
Noun	graun	NN	ground
Noun	sospen	NN	saucepan

---

[6]As the annotation at large still had to be done sentence by sentence by hand, any automatic insertion of an annotation tag was still checked and confirmed later in the process.

[7]Due to the limitations in Exmeralda's automatic tagging, capitalized and non-capitalized versions of the same item had to be tagged separately. Thus, for items frequently appearing in sentence-initial position, the capitalized version was included.

Noun	dewel	NN	devil
Noun	ples	NN	place
Noun	abus	NN	meat
Noun	muruk	NN	cassowary
Verb	kirap	VB	get up
Verb	kisim	VB	get
Verb	stap[8]	VB	
Verb	go	VB	go
Verb	wasim	VB	wash
Verb	kam	VB	come
Verb	tokim/tok/toktok	VB	
Verb	wokobaut	VB	walk
Verb	kamap	VB	come up
Verb	wokim	VB	work
Verb	givim	VB	give
Verb	pinis	VB	
Verb	karim	VB	carry
Verb	putim	VB	
Verb	tanim	VB	turn
Pronouns	em/Em	PRP	
Pronouns	yupela/Yupela; yu/Yu	PRP	you
Pronouns	mipela; mitupela; yumi	PRP	we
Pronouns	mi/Mi	PRP	I
Adjectives	lapun	PRP	old
Adverbs	nau/Nau	ADV	now
Adverbs	gen	ADV	again
Adverbs	bek	ADV	back
Determiners	dispela	DT	this
Determiners	tupela	DT	both
Interjections	yes/Yes	UH	yes

---

[8]Given that *stap* occurred frequently as both the TPI equivalent of *to be* and as progressive marker, its annotation was not completed automatically. For similar reasons, the following were also not automatically completed: the annotation of *tokim/tok/toktok* due to frequently appearing in the sense of *say* or *tell*; the annotation of *pinis* due to appearing frequently as both *finish* and as a completive marker; the annotation of *putim* due to both appearing as an equivalent of *put (down)*; the annotation of *em*, which does not indicate grammatical gender, whereas the third person singular pronouns in English do.

Interjections	orait/Orait	UH	allright
Prepositions	bilong	PREP-P	of
Other	i	PM	PM
Other	bin	PSM	PST
Other	bai	PST	FUT

Table 5.3: Auto tag list for annotation

5.2 Process of analysis

With the corpus ready for analysis, the individual innovative constructions to be surveyed had to be chosen. The challenge therein was twofold: first, as was established above, creativity and innovation are processes, and from recordings, we can only observe outcomes. In other words, the available data is linguistic shows the outcome of the contact situation Tok Pisin emerged in and thus the species, in Mufwene's terminology, that emerged as victorious in the evolutionary cycle. It does not necessarily show all other species contending for the same function, i.e. all the other features in the feature pool available to speakers of the emerging contact variety, but only the forms which eventually became grammaticalized, whether or not they arose through form-function reanalysis or other means such as direct borrowing.

Traces of creativity and innovation, then, were identified in two ways: first, through variation of forms fulfilling the same grammatical function, which indicates a competition between forms in the sense of Evolutionary Linguistics. For instance, the fact that Tok Pisin has three relativizers - *ia*, *we* and *husait* was a possible indication that they represent different creative solutions to the same communicative problem - i.e. that of expressing relativization. Secondly, traces of creativity and innovation may be found through comparison of functional items within Tok Pisin which have equivalents in either form or function, but not both, in contact languages whose main lexifier was English. For such tracing, the Atlas of Pidgin and Creole Structures offered valuable comparison data. There is no other English-lexified contact language, for instance, with the construction of *wantaim*, nor is there one where a temporal construction is adopted to serve as a comitative/instrumental preposition. Similarly, the form-function combination of *yet* as a reflexive marker all of the other contact languages adopt a form of English *self* as the reflexive marker or reflexive pronominal suffix, Tok Pisin opts for *yet*, a combination not attested elsewhere

among English-lexified contact languages.

In a survey of the grammatical features in Tok Pisin, such traces of innovation and competition could be found for the following features:

1. Nominal plural marker *ol*
2. Prepositional system, in particular the comitative/instrumental preposition *wantaim*
3. Adjectival suffix *-pela*
4. Transitive suffix *-im*
5. Predicate marker *i*
6. Coordination (cumulative, adversative and disjunctive conjunctions *na*, *tasol* and *o*)
7. *maski*
8. Verbal tense-aspect-mood markers *bin*, *pinis*, *i stap*, *wok long*, *kirap*, *kamap*, *save*, *inap*, *ken*, *traim*
9. Passive
10. Relativizers *ia*, *we* and *husait*
11. Complementizers *(bi)long*, *olsem*, *na*, *we* and zero
12. Reflexive *yet*

Including the initial survey, the analysis as a whole was carried out in seven steps:

1. Establish traces of innovation and competition to identify features of interest
2. Compare the features surveyed to closely related languages (i.e., Bislama and Solomon Islands Pijin)
3. Establish the original meaning/function of the constructing, should it serve multiple functions in modern Tok Pisin
4. Determine structural demand/supply, functional demand/supply and the origin of the construction as well as its path(s) of grammaticalization
5. Identify functionally competing constructions and their origin and path(s) of grammaticalization
6. Analyse factors which gave the "winning" construction an advantage and factors which gave competing constructions a disadvantage

The glossing in this study is mostly based on the Leipzig Glossing Rules, with some adjustments and extensions. For a full list of the glosses used, see table 8 in the appendix.

6 Analysis

6.1 Nominal plurality: *ol*

I will start my analysis of the origins of various constructions in Tok Pisin with *ol*. In the current varieties of the language, it serves as both a plural marker and a third person plural pronoun, as illustrated by example 3 (Mühlhäusler 2003, 71):

(3) Taim **ol** misinare i   kamap long Bismarck, **ol** kanaka i
    When PL missionary PM arrive at   Bismarck PL native PM
    no   lukim waitman   yet, nau **ol** i   kolim **ol** pikinini,
    NEG see   white-man yet, then they PM call   PL children,
    'Father'.
    'Father'.
    When the missionaries arrived at Bismarck, the natives had not seen white men yet, they tenn called the children 'Father'.

Prenominal plural markers are not necessarily rare among English-lexified creoles, as shown by the APiCS data below:

Language	Type of plural marking	Options
Early Sranan	variable (human/inanimate)	preceding marker, reduplication
Sranan	variable (human/inanimate)	prenominal marker
Saramaccan	variable (human/inanimate)	prenominal marker
Nengee	variable (human/inanimate)	prenominal marker
Creolese	variable (human/inanimate)	reduplication/postnominal marker
Trinidad English Creole	variable (human/inanimate)	prenominal/postnominal marker, plural suffix)
Vincentian Creole	variable (human/inanimate)	postnominal marker, plural suffix
Jamaican	variable (human/inanimate)	postnominal marker, reduplication
Belizean	variable (human/inanimate)	prenominal/postnominal marker, plural suffix

San Andreas Creole English	variable (human/inanimate)	postnominal marker, plural suffix
Nicaraguan Creole English	variable (human/inanimate)	postnominal marker, plural suffix
Bahamian Creole	variable (human/inanimate)	plural suffix, postnominal marker, stem change
Gullah	variable (human/inanimate)	stem change, postnominal marker, plural suffix
African American English	invariant	plural suffix, stem change
Krio	variable (human/inanimate)	postnominal marker, reduplication
Ghanaian Pidgin English	variable (human/inanimate)	plural suffix, reduplication, stem change
Nigerian Pidgin	variable (human/inanimate)	postnominal marker, reduplication, stem change, plural suffix
Cameroon Pidgin English	variable (human/inanimate)	postnominal marker, reduplication
Pichi	variable (human/inanimate)	postnominal marker, plural suffix
Chinese Pidgin English	variable (human)	plural suffix, stem change
Singlish	variable (human/inanimate)	plural suffix, stem change
Tok Pisin	variable (human/inanimate)	prenominal marker
Bislama	invariant	prenominal marker, plural suffix
Norf'k	variable (human/inanimate)	prenominal marker, plural suffix
Kriol	variable (human/inanimate)	prenominal marker, reduplication, plural suffix
Hawai'i Creole	variable (human/inanimate)	plural suffix, stem change

Table 6.1: Nominal plural marking systems in the 26 English-lexified languages in the APiCS data

As noted above, Bislama also predominantly uses prenominal plural marking. As Meyerhoff (2013) notes, "plural nouns can be either *ol* + N or *olgeta* + N" while "marking with suffix *-s* from English is marginal." The example given is reproduced below:

(4) ol nes oli bisi tumas
     PL nurse AGR busy too.much
     The nurses were all too busy.

Solomon Islands Pijin, on the other hand, has no formal plural marking. To quote Huebner (1979, 38), "in Pijin, many times when the context is clear, number (singular or plural) is not marked on the noun", and even when it is necessary by context to do so, there is no general plural marker. Instead, Pijin opts for more specific plural marking with *wanfala*, *tufala*, *samfala*, *olketa*, or *evri* (Huebner and Horoi 1979, 38)

(5) Mi   lukim wanfala/tufala/samfala/olketa/evri pigpig.
     1SG see   NUM/NUM/DT/DT/DT   pig
     I saw one/two/some/some/all pig(s).

In order to explain why the plural marker in Tok Pisin and at least one of its related languages is still worth considering a special case, it is necessary to look at the diachronic development of both *ol* and *wantaim*. As detailed in chapter 4.1 above, Mühlhäusler identifies four stages in the development of Tok Pisin: the jargon stage, the stabilisation stage and the early and late expansion stage, respectively. In these, pluralization developed as follows: in the jargon stage, no formal plural marker exists. Plural could be expressed through purely semantic means, such as saying *plenti road* for 'many roads'. During the stabilisation stage, *ol* becomes used as a plural pronoun, but not yet a plural marker. Just as in the jargon stage, formal plural marker exists. In the early expansion stage, nominal plurality is expressed for animate nominals, especially in subject position, by preceding *ol*. In other words, we start seeing constructions such as *ol man* meaning not 'all men', but plural 'men'. In the late expansion stage, nominal plurality is expressed for animate as well as inanimate and abstract entities and in subject position as well as oblique case, etc., by preceding *ol*, like in the example at the beginning of this chapter.

In none of these stages, either formal plural marking by suffixation or any other morphological means was productive. In theory, there are several possible explanations for Tok Pisin not adopting morphological plural marking in favour of expanding the functions of *ol*: the input could have been absent or insufficient. There could have been interfering input from other substrates, for instance the local Papuan languages. The input could have been incompatible with the emerging system. The first explanation we might discard in its strong form (input being entirely absent) by referring to the fossilized lexical plural forms such as *anis* and *binen* presented above. However, these two fossilized forms do not provide strong evidence for a widespread input of the plural *–s* suffix. After all, both ants and bees usually occur as not a single animal, but rather a whole group. It is therefore doubtful that the distinction between the

English singular and plural forms would have been transparent.

Mühlhäusler and Wurm (1985: 114) do note though, "studies of the linguistic input in the formative years of Tok Pisin have shown that plural marking by means of the *-s* affix was a widespread feature of Pacific English Foreigner Talk." (Wurm and Mühlhäusler 1985, 114). But could the input have been insufficient instead? We will return to that idea below.

In terms of interfering input from Papuan languages, we have to consider that they exhibit two strategies. Some have no nominal plural marking at all, and could have provided a negative pattern for plural marking in the emerging Tok Pisin. In other words, they could have supplied the framework for a language system without *morphological* plural marking. The other set of Papuan languages has a very complex system of noun classes, with irregular marking of both dual and plural. Due to their complexity, they could not have been a model for Tok Pisin, as their systems lack the criterion of transparency (which we will return to below).

Next, let us examine what factors could have made morphological plural marking an unsuitable form in the emerging language structures of Tok Pisin. Mühlhäusler (1985, 115) makes two arguments as to this point:

> Affixation to signal plurality of nouns, as in English or German , clearly conforms to the principle that plural forms should, from the point of view of the ease of perception, be longer than singular ones. However, affixes are less accessible than free forms. Since the optimalisation of perception characterises the early development of a pidgin, one would not expect affixes to be borrowed until the pidgin is structurally and functionally comparable to a first language. It is for this reason that the free form *ol* [...] emerges in Tok Pisin, and that neither English *-s* nor German *-en* had a good chance of being borrowed.
>
> A second important argument [...] is the following: if a pidgin develops plural marking, it will appear first in the most natural environment (animates in subject position) and then spread to less natural ones. We find that the lexical items containing English or German plural affixes do not provide a favourable environment for the spread of a plural rule.

In addition, there was no systemic precedent for morphological marking of any kind. Neither verbal nor nominal or adjectival inflection was present,[1] with functions such as pluralization, comparison or possession either being absent or being expressed by syntactic means, such as the possessive preposition *bilong.*

---

[1]Compare also Landtman (1918), who notes that "As a rule only the simplest indicative form of the English verb is used" and that it is only "in exceptional cases [that] such a form as "I says" may occur"

The introduction of inflectional plural marking would therefore have meant the introduction of inflectional morphology as a concept and process, making plural inflection highly marked. As, for instance, DeGraff (2001, 509; 514) and Muysken (1981) have argued, it is the unmarked option which usually prevails in creole grammar and often in koenization, as well. This is in line with the second morphological constraint on structural diffusion Winford (2009, 96) proposes:

> The greater the degree of transparency of a morpheme, the greater the likelihood of its diffusion. By contrast, the more opaque (complex, bound, phonologically reduced) a morpheme is, the less likely it is to be borrowed.

An inflectional morpheme in a language that has no inflectional morphology certainly does not fulfil the criterion of transparency. This would put it at a disadvantage when competing with an isolated form, which is easier to parse. In addition, both *anis* and *binen* are likely to have entered Tok Pisin as plurals, since both are usually not encountered as single entities, and therefore far less likely to be referred to as such in language use. This relates back to the idea of insufficient input due to to the lack of transparency Winford describes, as neither *-n* nor *-s* in these items are likely to have been recognized by speakers as pluralising morphemes.

It was, of course, not entirely impossible for the emerging Tok Pisin to become a fusional language, or to at least adopt some morphological marking. As Matras (2007, 40) notes, "a number of languages show signs of movement between morphological types". However, he also remarks that "none of these developments seem to follow any predictable structural path, and the only common denominator is an accommodation to the patterns of a socially dominant contact language". While the borrowing of inflectional morphology is something that by common wisdom is rare in language contact, it has been observed (see Meakins 2011, 87), and shown to happen on a cline from full retention to full loss, rather than identically across contact situations and that it is particularly inherent inflections such as plural marking which may be retained (Roberts and Bresnan 2008). In Field's (2002, 38) borrowing hierarchy, fusional affixes form the very end, being borrowed only after content items, function words and agglutinating affixes. Heath (1978, 105ff.) suggests that factors such as morpheme syllabicity, the sharpness of boundaries between morphemes, unifunctionality of morphemes and the categorical clarity of morphemes could disfavour the borrowing of inflectional morphology. The first two, at the very least, would disfavour a borrowing of English plural *-s* in cases such as *anis*, where they are not perceived as individual morphemes,

but as part of the lexical root.

Mühlhäusler further states that "[t]he question remains, why reduplication was not borrowed from Tolai as a plural-signalling device [...]" and concludes that "to this I do not have an answer." To understand, let us take a look at how reduplication works in Tolai. As Mosel (1980, 104) points out, "[n]ouns that are not derived from verbs by the nominalisation affix /ni-, -in-, -un-/ can be reduplicated in order to express the plurality of the object referred to by the simplex". Her examples include *barmana* 'young man' to *barbarmana* 'young men', *gunan* 'village' to *gunagunan* 'villages' and *uma* 'garden' to *umauma* 'gardens'. A very similar type of noun reduplication exists in Tok Pisin as well, but did not become productive as a plural marker. As Mühlhäusler (1979, 416f.) points out, "[w]ord level reduplication of noun bases, and occasionally compound nominals, is used to express the idea that a considerable number of what is referred to by N are involved", as for example in *hulhul* 'many holes' or *long naitnait* 'every night'. The question, however, is whether this indicates grammatical plural or a multitude on a lexical level. Reduplication in this form seems to be unable to express a lower plural, e.g. just two entities. That is to say, *naitnait* refers to every single night, not just two nights. Many holes are not just two holes, but a significant number of holes. (Wurm and Mühlhäusler 1985, 115)

I suggest that one factor in the emergence of *ol* as a nominal plural marker was the fact that reduplication was already in place as a means of lexical innovation. In derivation, it often expressed other functions than a grammatical plural, including continuous, reciprocal and repetitive actions in verbs such as *lukluk* and *toktok* or variety in nouns such as *kala-kala*. While these functions are related to plurality - *kala-kala*, as a variety of colours, for instance, by necessity indicates more than one colour - they operate on a lexical level instead of a grammatical level.[2] This would indicate that reduplication served mainly a derivational function, not an inflectional one. While this does not preclude it being imported as or being later employed as an inflectional function as well, it is, in addition to the points laid out above, another factor weighing in in favour of another solution - granting a small advantage to other variants such as *ol* in the feature pool.

Aside from the structural factors explored above, the point of emergence for the *ol* plural marking is also significant for several reasons. According to Mühlhäusler's four stage development model for Tok Pisin, the plural marker emerged during the early expansion stage, after stabilization. This early ex-

---

[2] Also note the argument in Hall (1943b, 194) that reduplication served an onomatopoeic function.

pansion stage can be roughly dated to 1880 - 1914. During this stage, input of English from L1 speakers was limited because in 1884, the north-eastern part of New Guinea was annexed by the German Empire. Contact with speakers of English on one hand and speakers of other Melanesian pidgins on the other was cut off as the German colonial administration stopped the labour trade with other plantations and pulled native New Guineans back to the main islands (Romaine 1992a, 37). While German never came to be widely used by the local population, heavy investment into infrastructure by the German colonial supervision led to the spread of Tok Pisin to rural areas. Being cut off from its main lexifier led to internal developments in Tok Pisin relying on existing structures. In this case, this meant adopting a personal pronoun as plural marker. In other words, once the categorial demand for plural marking was pressing, formal supply of the source language(s) was either absent or unsuitable. So even if the stabilization stage had led to a system that was stable enough to accept morphological plural marking, the input change from L1 English to L2 English would have made such a borrowing more unlikely[3]. It is also noteworthy that, as Smith (2002b, 65) reports, "recently the -*s* suffix has become increasingly used to signal plurality as Tok Pisin and English come into increasingly frequent contact". This lends credence to the argument that morphological plural marking is not entirely incompatible with Tok Pisin, but timing and the contact situation played a role.

There is a further caveat, however, in that even in contemporary Tok Pisin, plural marking with *ol* is not obligatory. As Verhaar (1995, 346ff.) notes, rather than just providing a grammatical function of pluralization, *ol* can also be seen as marking the following noun as a collective. On the other hand, semantic plurality does not automatically require grammatical marking with *ol*. This was likely another factor that hindered a one-to-one replication of English plural marking, which is obligatory. If the categorial demand was for a context-dependent collective marking rather than a context-independent plural marking based on semantic plurality, English obligatory plural marking would have been even less suitable.

---

[3]Note, however, that the fact that closely-related Bislama, which was not cut off from English entirely during that time, also adopted *ol* as plural marker going back at least as far as 1913, see Crowley (1998, 90) Therefore, the contact situation cannot have been the only relevant factor.

6.2 Comitative and instrumental: *wantaim*

Wantaim is the comitative and instrumental preposition in Tok Pisin, meaning it is used to convey both accompaniment as in *walk with someone* and an instrument or means by which an action is accomplished, as in *she opened the box with scissors.* Examples 6 and 7 below show it being used in both of these functions in Tok Pisin[4]:

(6) Mangi,   bebi boi, ol   i   go **wantaim** papa, na   liklik
Young-people, baby boy, 3PL PM go PREP-C father CONJ small
meri   olsem ol   i   go **wantaim** ol mama.
woman likewise 3PL PM go PREP-C PL mother.
The young people, the baby boy, they went with the father, and the small girls likewise went with the mothers. (ZC 03814)

(7) Karim **wantaim** bet   na   ol   wokobaut   nau,
Carry PREP-I   stretcher CONJ 3PL walk-around now,
wokobaut   i   go na   ol   singaut.
walk-around PM go CONJ 3PL call-out.
They carried him with a stretcher and walked around now, they walked around and they called out. (ZC 03209)

In these examples, modern Tok Pisin's *wantaim* seems functionally (or categorially) identical to English *with.* Structurally (or formally), however, *wantaim* seems to be adapted from English *one time.* Thus, its functions in modern Tok Pisin are not immediately apparent from the perspective of an English speaker. To her, it might seem peculiar that what looks like a temporal expression is used to serve a comitative or an instrumental function. One possible explanation for this phenomenon would be that the form *time* itself could have been borrowed into Tok Pisin with a wholly different meaning, or stem from a different source language that happens to have a similar form attached to a different meaning. However, this is not the case. As is made evident by the example below, aside from *wantaim, taim* is used exclusively with a temporal meaning in the language (Mühlhäusler 2003, 267; 274)[5]

(8) Mi stori ia   **pastaim**. Bai yumi   go lo   moning**taim**.
1SG story FOC first.   FUT 1PL-INC go PREP-L morning.
First, my story. We will go in the morning.[6]

---

[4]ZC = Z'graggen Corpus

[5]Glossing by source and adjusted for consistency within present paper, translation mine.

[6]In TP, both *lo* and *long* are forms of the same preposition with an abstract spatial meaning roughly equivalent to English *at.* At various developmental stages of Tok Pisin, forms of *long* can also be used with temporal expressions, equivalent to English *at night.*

Further indication that Tok Pisin's comitative/instrumental preposition is an unusual case is found by comparing the source of comitative and instrumental prepositions of the 26 English-lexified languages in the Atlas of Pidgin and Creole Structures Michaelis et al. (2013):

Language	Comitative	Instrumental	Origin
Early Sranan	*langa / nanga*	*langa / nanga*	English (*along*)
Sranan	*nanga*	*nanga*	English (*along*)
Saramaccan	*ku*	*ku*	Portugese (*com*)
Nengee	*anga*	*anga*	English (*along*)
Creolese	*wid*	*wid*	English (*with*)
Trinidad English Creole	n/a	n/a	n/a
Vincentian Creole	*wid*	*wid*	English (*with*)
Jamaican	*wid*	*wid*	English (*with*)
Belizean	*wid*	*wid*	English (*with*)
San Andreas Creole English	*wid*	*wid*	English (*with*)
Nicaraguan Creole English	*wid*	*wid*	English (*with*)
Bahamian Creole	*with*	*with*	English (*with*)
Gullah	*wid*	*wid*	English (*with*)
African American English	*with*	*with*	English (*with*)
Krio	*wit*	*wit*	English (*with*)
Ghanaian Pidgin English	*wit*	*wit*	English (*with*)
Nigerian Pidgin	*wit*	*wit*	English (*with*)
Cameroon Pidgin English	*wit*	*wit*	English (*with*)
Pichi	*wet*	*wet*	English (*with*)
Chinese Pidgin English	*long*	*with*	English (*along, with*)
Singlish	*with*	*with*	English (*with*)
Tok Pisin	*wantaim*	*long, wantaim*	English (*along, one time along*)
Bislama	*wid*	*wid*	English (*along, with*)
Norf'k	*lorng*	n/a	English (*along*)
Kriol	*wit*	*wit*	English (*with*)
Hawai'i Creole	*wid*	*wid*	English (*with*)

Table 6.2: Instrumental and comitative prepositions and their sources in the 25 English-lexified languages in the APiCS data

As shown in table 6.2 above, the common trend among English-lexified creoles

is to adopt English *with* for the comitative and instrumental, with *(a)long* being
the prime alternative choice. Even Bislama, which is so closely related to Tok
Pisin that some have called it dialects of the same language (Keesing 1988, 3),
uses *wid*. In addition, to the best of my knowledge, there is no other case in
which a temporal construction was grammaticalized to serve as a comitative or
instrumental preposition. As with *ol*, the borrowing and grammaticalization
process in the emergence of Tok Pisin seems to run counter to crosslinguistic
tendencies and expectations in pidgins and creoles. The question remains,
then, why and how speakers of Tok Pisin or its predecessors would have come
up with such an exotic construction as *wantaim*.

Modern Tok Pisin has three different simple prepositions - *bilong*, *long* and
wantaim - whose functions overlap to a certain extent, as table 6.3 shows:

Prep.	Function(s)	Examples
bilong	all purpose preposition; in modern TP denotes possession, purpose, origin, charateristic trait	*Han* **bilong** *mi em i doti* ('My hand is dirty', Steinbauer 1969, 43)
(a)long	mostly spatial relationships; used for English prepositions *in, on, at, to, from, by, about, because, for, during, with* (especially as instrumental)	*Em i bengim ka* **long** *rot* ('He smashed the car on the road', Steinbauer 1969, 23)
wantaim	corresponds to comitative/instrumental *with*	*Mi miksi muli wara* **wantaim** *ti* ('I mix lemon water with tea', Steinbauer 1969, 119)

Table 6.3: Prepositions and their functions in modern TP

As mentioned above, there are several issues concerning the emergence and
grammaticalization of *wantaim*. The first concerns its origin: how and why
was what appears to be a form of English *one time* adapted to serve as a
comitative/instrumental function in Tok Pisin? The second is concerned with
its status within the prepositional system. How and why was *wantaim* able
to assert itself as an independent preposition from the ubiquitous *long* while
other complex prepositions such as *arere long* have not?

A look at Bislama and Solomon Islands Pijin, which are the two most
closely related languages to Tok Pisin, shows that the development itself - or at
least its outcome - is unique to Tok Pisin. The comitative constructions found
in examples 9 (Bislama, Crowley 2004, 27) and 10 (Solomon Islands Pijin,
Jourdan and Keesing 1997, 407) are not etymologically related to *wantaim* at
all. [7]

---

[7]As one anonymous reviewer on Eberl 2019 remarks, Bislama *wetem* has been attested

(9)   Hem i   stap   slip **wetem** tufela.
       3SG PM PROG sleep PREP-C PRN

He lives with the two of them.[8]

(10)   Mi nao   mi faet **wet-em** olketa nomoa.
       Me TOP I   fight with-TR them just.

Me, I was just fighting with them.[9]

The answer to why Tok Pisin would use a seemingly outlandish construction such as *wantaim* lies in the very first of the functional stages in the grammaticalization of *wantaim*, which was the adverbial construction *one time along*. To quote Mühlhäusler, "[w]antaim, originally an adverb meaning at the same time, is frequently used as a preposition translating the concepts of *with, together with* and *with the use of*" (Wurm and Mühlhäusler 1985, 367). However, while this tells us where it originally came from, it does not yet explain the further steps in its development, nor does it explain the motivation behind the reanalysis of a temporal adverbial as the comitative. In other words, we lack the information of "how". I propose that the diachronic development of *wantaim* through the emergence of Tok Pisin can be traced as follows:

- Stage 1: Temporal adverbial

- Stage 2: Temporal adverbial plus comitative preposition

- Stage 3: (Temporal adverbial plus) comitative preposition (plus instrumental preposition)

- Stage 4: Comitative preposition plus instrumental preposition

In regards to stage 2, early attested usage of prepositional *wantaim* and *one time along* is strictly comitative[10].

(11)   **One time along** taro.
     PREP-C   taro.

With taro. (Mead 1931, 42)

---

     since 1914, being cited in an example *Me me go widim you* in (Crowley 1998, 103). Later re-spelled *wetem*, the reviewer notes that it might be a fusion of English *with* and the transitivity marker *-im*. This sounds very plausible to me. One can only speculate about whether *wantaim* would have received an additional *-im* as well, had it not already possessed the ending.

[8]Glossing and emphasis mine, translation by original source.

[9]Glossing and translation by source cited. TOP = topical marker, TR = transitive marker.

[10]For examples 13 through 17, glossing and translation mine.

(12) Long   nait mi wokabaut   **wantaim** sutlam.
PREP-L night 1SG walk-around PREP-C   flashlight.

At night, I walk around with a flashlight. (Steinbauer 1969, 193)

(13) Yu   dring solmarasin **wantaim** wara.
2SG drink epsom-salts PREP-C   water.

You drink epsom-salts with water. (Steinbauer 1969, 183)

Mihalic (Mihalic 1957, 159) still lists *long* as part of the construction, though it seems to have become optional by the time his source data were collected. The entry in his dictionary reads "with: *wantaim, wantaim long*; to work with him = *wok wantaim (long) em*". Meanwhile, the instrumental function was still expressed almost exclusively by *long*:

(14) Mi   kisim rais **long**   skel.
1SG take   rice PREP-I scale.

I take rice with the scale. (Steinbauer 1969, 177)

(15) Katim pepa **long**   sisis.
Cut   paper PREP-I scissors.

Cut the paper with the scissors. (Steinbauer 1969, 175)

At the same stage, the form *wantaim* served several functional uses, as the following excerpt from Mihalic's dictionary (Mihalic 1957, 159) shows:

wantaim

1) one time, once

Mi mekim wantaim tasol. = I did it only once.

2) at the same time, with, together, and

go wantaim = to go along with, to accompany

kisim pensil wantaim pepa = to take pencil and paper

pasim wantaim = to tie together

tupela wantaim = both together

wok wantaim = work together, to work at the same time

3) to express similarity or equality

Tupela i-strong wantaim. = The two are of equal strength.

This is an expression used to denote a tied score in a game, or of a battle evenly fought.

Similar to Mihalic's entry, Hall's (1943a, 124) vocabulary section records *wantaim long* as a prepositional phrase with the meaning of 'with'. In addition, it

also contains an entry for *wantaim* on its own, listing it both as an adverbial meaning "(at) one time, at the same time, together" and as a quasi-preposition meaning 'with'. As in Mihalic's dictionary, there is no indication of an instrumental function yet.

The third stage saw the expansion of *wantaim* to include the instrumental function, as in the examples below[11]:

(16) Ol nes i ken samapim maus bilong yu **wantaim** string.
PL nurse PM can sew mouth PREP-P 2SG PREP-I string.

The nurses can sew up your mouth with string. (Wurm and Mühlhäusler 1985, 368)

(17) Ol i sutim disla pig ia **wantaim** disla ol spia bilong
3PL PM shoot DT pig EMPH PREP-I DT PL spear PREP-P
ol.
3PL.

They shot this pig with these spears of theirs. Smith (2002a)

Verhaar (1995, 251) notes that while instrumental *wantaim* is still rare, Tok Pisin seems to mirror the development from comitative to instrumental that occurred in many languages. This development is further confirmed by Siegel and Smith (2013; additional online dataset), who note that the usage of instrumental *wantaim* is increasing in present-day Tok Pisin. They claim that "the general preposition *long* is most commonly used for instrumental and *wantaim* for comitative. But the use of *wantaim* for instrumental is increasing as a result of the effect of English." Meanwhile, the temporal adverbial *wantaim* was mostly replaced by *wanpela taim*, as in example 18 (Siegel and Smith 2013, 219) [12].

(18) **Wanpela taim** em laik go long bus [...]
NUM time 3SG like go PREP bush [...]

Once he wanted to go to the bush [...]

The origin of *wantaim*

Having established that the comitative use of *wantaim* predates the instrumental use, it is evident that in order to determine the origins of prepositional

---

[11]For example 16, glossing mine and translation from original source. For example 17, both from original source, with glossing adjusted.

[12]Glossing mine, translation by original source.

wantaim, the focus needs to be placed on the comitative on the functional side. In regard to the formal side, I have mentioned above that the original form of the construction was *one time along.* This form was by no means as unique as *wantaim* is today, but in fact was modelled after a productive pattern.

At the time of *wantaim*'s origin, *(a)long* was already in use as a preposition in addition to the earlier all-purpose preposition *bilong. One time along* was modelled on a pattern in which *along* combined with (mostly spatial) adverbs to form complex prepositions. *Long* formed the basic prepositional element and X further specified its reading, as in *arere long X* 'alongside X' or *aninit long X* 'underneath X'. The question, then, is what exactly *one time* specified in the construction of *one time along.* According to Mühlhäusler 1985, 367, *wantaim* was originally an adverb meaning 'at one time'. Given the structure of adverbs and prepositions in Tok Pisin - the first ones occuring as simple lexemes, while the latter are, with the exceptions above, formed with *long* - it seems likely that *one time* was the adverbial form and *one time along* arose from it as a preposition in the process described above.

Essentially, there was a spatial concept – *(a)long* – which was expanded with a temporal concept *one time.* As an adverbial construction, *wantaim* signified 'at one time', as in a hypothetical English sentence *at one time, I was hanging upside down from a tree.* The prepositional form does, however, by its nature, require a prepositional complement with whom the point in time is shared. The meaning of the combined *one time (a)long* becomes 'one point in time at X', as in "One time, in the garden, I was reading". [13]

From there, the step to the comitative is, cognitively, not that far-fetched. Under the assumption that the cognitive basis of the comitative is 'same time plus same space', it is easy to see how a reanalysed reading of *one time along* became not 'at one time', but 'same time at' or 'same time space'. In addition, *one* may easily have been read and reanalysed as applying to both elements, and been understood as 'one place, one time', or 'same time, same space'. The suggested development is summarized below:

$$X + along \Rightarrow \text{'at X'}$$
$$\Downarrow$$
$$\textit{one time along} \text{ 'at one time'}$$
$$\Downarrow$$
$$\textit{one time along} \text{ 'same time at'}$$
$$\Downarrow$$

---
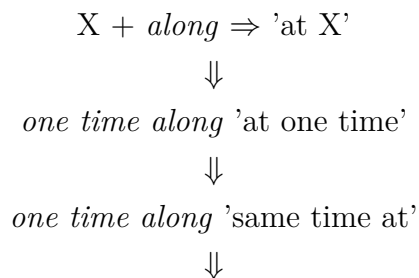
[13]Claims that *one time along* would have been parsed as 'at one time' seem unlikely, given that *(a)long* is always followed by the prepositional complement.

$$\textit{one time along} \text{ 'same time + space'}$$
$$\Downarrow$$
$$\textit{one time along} \text{ 'same time, same space'}$$
$$\Downarrow$$
$$(\text{equal (time+space)})$$
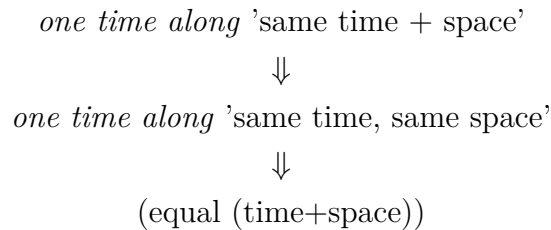
While grammaticalization of the comitative from a temporal source may be typologically uncommon, it is not unheard of. A precedent may be found in the Latin adverb *insimul* 'together' (Glare et al. 1982, 925), the root of the adverbs *insieme* in Italian 'together, at the same time' and *ensemble* 'together' in French. While they are not comitative prepositions, these adverbs express the same concept: being in the same or very near space at the same time. *Insimul* in turn derives from *in- + simul*, the latter of which means 'at the same time', 'together' or 'as soon as' and can serve as an adverbial, a conjunction or a preposition with the ablative (Glare et al. 1982, 1766). It thus serves at least two functions that *wantaim* also serves: that of a comitative preposition and an adverb, as well as sharing a temporal and comitative aspect. The grammaticalization path is not quite the same, however, given that *simul* is derived from *similis* 'similar, like' or 'same'. This actually details the third aspect in the construction of *one* plus *time* plus *along*: that of similarity, identity or equality. In the case of Latin, we therefore have a precedent that contains two parts of the trifold concept expressed in *wantaim*: *similis* for equality and *simul* for the temporal aspect. It may therefore seem tempting to ascribe some sort of spatial notion to the *in-* prefix in *insumul*. And indeed, Latin *in-* could convey a meaning of 'in, inside', but for the most part, only when affixed to verbs. On adjectives, it mainly reversed the meaning of the adjective, as in *gnarus* 'knowledgeable', but *ingnarus* 'ignorant'.

Furthermore, there is another link in the opposite direction. Heine and Kuteva (2002, 89f.) report several instances of the comitative > temporal grammaticalization path, among them German instrumental/comitative *mit* > temporal *mit* as in *Mit achtzehn Jahren begann sie ein neues Leben* 'At 18, she began a new life'. They note that "while there appears to be sufficient evidence to support this grammaticalization, more research is required on the conceptual basis of the process", suggesting that it is conceivable that the pathway leads from comitative to instrumental to temporal rather than directly from comitative to temporal. As we have seen, however, in Tok Pisin, comitative *wantaim* predates instrumental *wantaim*. Therefore, the path from temporal to comitative might be a different path - if no less complex - than from comitative to temporal, or a path with a different direction.

The notion of *one* or *wan* expressing equality is strengthened by two observations from Mihalic's dictionary. The first is the third section of the entry under *wantaim*, which chronicles the following function [14] (Mihalic 1957, 159):

3) to express similarity or equality

Tupela i-strong wantaim. = The two are of equal strength.

This is an expression used to denote a tied score in a game, or of a battle evenly fought.

Secondly, there are a number of parallel constructions in the lexicon of Tok Pisin, all expressing the idea of being equal or identical (Mihalic 1957, 258):

same, *wankain*

at the same time = *wantaim*

from the same village = *wanples*

from the same country = *wantok*

living in the same house = *wanhaus*

of the same age, class = *wanlain*

of the same kind = *wankain*

of the same name = *wannem*

of the same nationality = *wantok*

of the same size = *wanmak*

of the same tribe = *wanpisin*

speaking the same language = *wantok*

The productivity of the pattern *wan* + X is also the first of several factors that contribute to answering the second question above: how was *wantaim* able to assert itself as an independent, simple preposition over other complex prepositions. Of the latter, quite a few exist (Verhaar 1995, 236): "*aninit long* 'below, under, underneath'; *antap long* 'on, on top of, over, above'; *arere long* 'alongside (of)'; *ausait long* 'outside'; *baksait long* 'behind, at the back of; *bihain long* 'after'; *bipo long* 'before [of time]'; *inap long* 'until, as far as'; *insait long* 'inside'; *klostu long* 'near, close to'; *namel long* 'between'; *paslain long* 'before [of place], in front of; and *raun long* 'around'". In none of these, *wan* + X would have played a role. These prepositions lacked the additional

---

[14]While this function is absent from the samples drawn from the Z'graggen corpus, I am assured by an anonymous reviewer on Eberl (2019) that they witnessed this use "quite often during fieldwork in Papua New Guinea".

productivity boost of the construction.

The major structural change from its early construction *one time along* to *wantaim* is the loss of the accompanying *long*, which is still obligatory for most of the other complex prepositions. In all of the constructions with *wan* cited in the dictionary excerpt above, however, *long* does not appear, for the simple reason that these are not prepositional constructions. The *long* in *wantaim long* has, therefore, been lost in a process of analogization of the first part of the construction.

Another contributing factor was the frequency of *wantaim*, which is significantly higher than that of other constituents of complex prepositions, as this data from the Z'graggen corpus shows:

Preposition	meaning	Occurrences(total)	Occurrences(per million words)
antap	on top of	32136	3785
wantaim	with	29997	3533
insait	inside of	15655	1843
bipo	before	14012	1650
inap	until	12591	1483
bihain	after	9668	1139
klostu	close to	8445	995
raun	around	3541	417
arere	next to	2929	345
aninit	underneath of	2222	262
baksait	behind of	1770	208

Table 6.4: Number of occurrences of the first constituent of prepositional constructions

There is one caveat to observe here, however. Given the fact that the corpus is not POS-tagged, these numbers include prepositional as well as adjectival and adverbial uses of these lexemes. Further examination would therefore be necessary to validate this argument.

The third factor I posit is the possibility that as its grammaticalization continued, the link between *wantaim* and the spatial dimension was weakened, which further encouraged the drop of *long*. Since the comitative has a spatial dimension alongside a temporal one, but is not exclusively spatial, the cognitive link between *long* and *wantaim* was weaker than, for instance, the link between

aninit or *arere* and *long*. This possible factor is further pursued in the data below. As an exploratory analysis, in the same corpus as above, *aninit* and *arere* - which both have exclusively spatial functions - are followed by *long* roughly 75% of the time, while *wantaim* is only followed by long in about 2% percent of all cases, as the numbers in table 6.5 show. *Bipo*, which serves a primarily temporal function, is rarely followed by *long* in the data as well.

Preposition	Occurrences followed by long	Primary function
wantaim	619 (2.06%)	comitative
bipo	817 (5.83%)	temporal
aninit	1642 (73.90%)	spatial
arere	2112 (72.11%)	spatial

Table 6.5: Number and percentages of occurrence of complex prepositions

However, this data is not conclusive due to the fact that it includes both adverbial and prepositional occurrences of the forms in question. Therefore, another sample of 250 occurrences was drawn for *bipo* and *bihain*, which can serve as both temporal and spatial adverbs and prepositions, as well as *arere* and *aninit*, which can only serve as spatial adverbs and prepositions. Table 6.6 and figure 6.1 below summarize and illustrate, respectively, the occurrence of their various functions in the sample with or without *long*.

-	ADV-T	ADV-L	PREP-L	PREP-T
bihain (+*long*)	0	0	27	5
bihain (-*long*)	82	30	0	105
bipo (+*long*)	0	0	2	6
bipo (-*long*)	161	1	3	74
aninit (+*long*)	0	0	193	0
aninit (-*long*)	0	56	0	0
arere (+*long*)	0	0	139	0
arere (-*long*)	0	103	2	0

Table 6.6: Occurrences of different functions of *bihain, bipo, arere* and *aninit* with and without *long*

As is immediately evident, neither adverbial *bihain* and *bipo* nor adverbial *arere* and *aninit* ever occur with *long*, which means we can immediately discard them for the present analysis. Secondly, spatial prepositions seem to show a strong
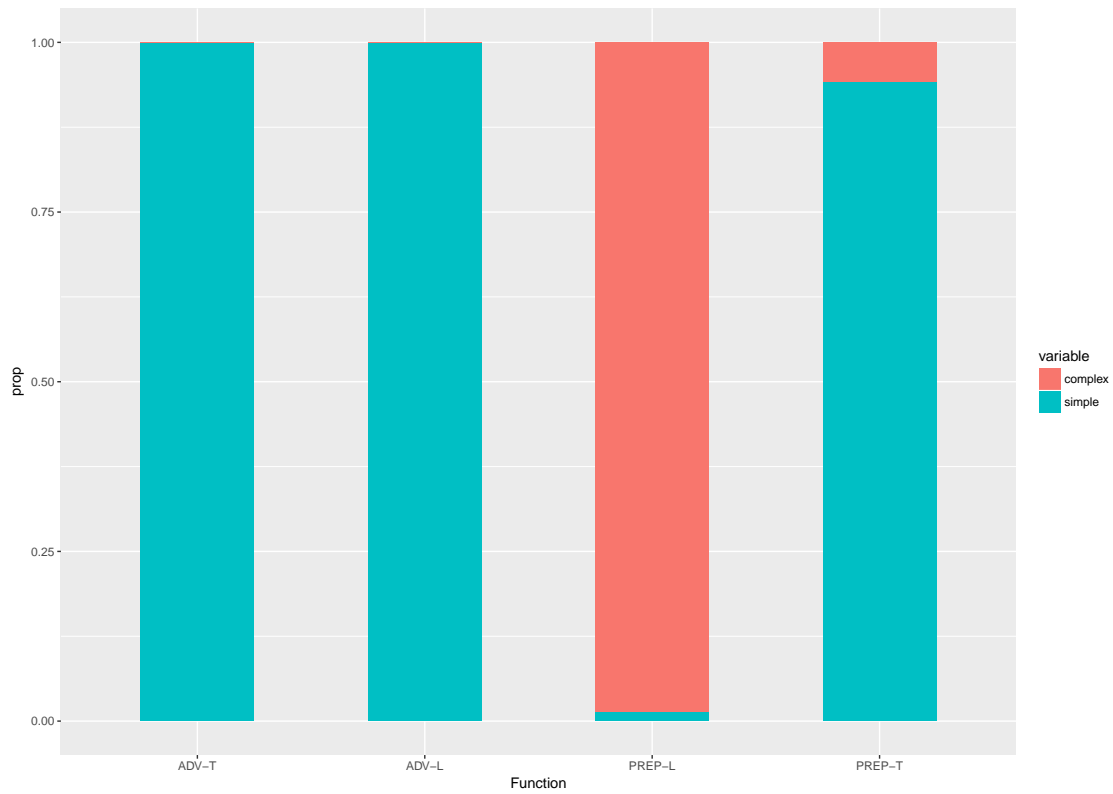
Figure 6.1: Occurrences of different functions of *bihain, bipo, arere* and *aninit* with and without *long*

tendency within this sample of *bihain* and *bipo* to require *long*, while temporal prepositions seem to strongly favour discarding it, which would speak in favour of a link between the spatial dimension and *long*.

Form	PREP-L	PREP-T
bihain (+*long*)	48	17
bihain (-*long*)	-	185
bipo (+*long*)	5	42
bipo (-*long*)	3	200
aninit (+*long*)	250	
aninit (-*long*)	-	
arere (+*long*)	247	
arere (-*long*)	3	

Table 6.7: Occurrences of different functions of *bihain, bipo, arere* and *aninit* with and without *long*
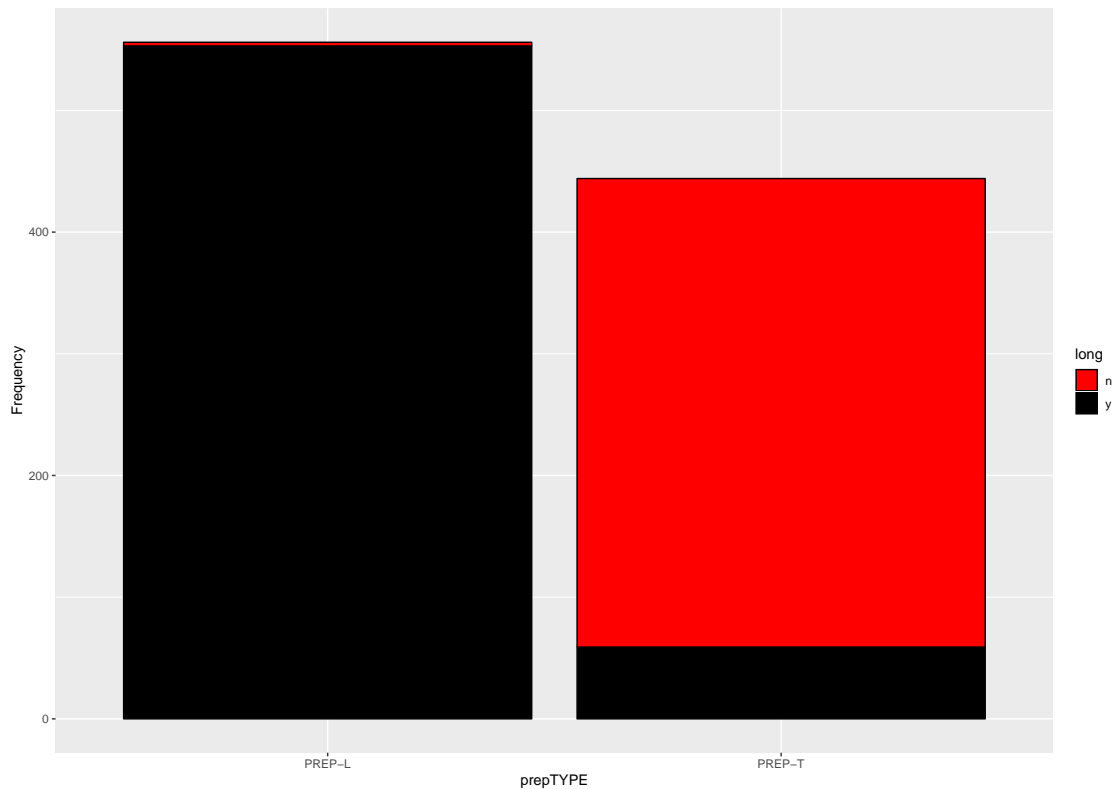
Figure 6.2: Occurrences of different functions of *bihain, bipo, arere* and *aninit* with and without *long*

To substantiate this further, the subset of prepositional occurrences of the previous sample was extracted. Further prepositional occurrences were then sampled until 250 purely prepositional occurrences for each were available for analysis as to their prepositional type (spatial or temporal), the presence or absence of *long* and their prepositional object. The data is summarized in table 6.7 and figure 6.2. As is immediately evident, the vast majority of spatial prepositions occurs with *long*, whereas the majority of temporal prepositions does not. Temporal prepositions do, however, occur with *long* more often than spatial prepositions occur without it. The phi-coefficient for the association between the presence/absence of *long* and the type of preposition (spatial or temporal) is 0.879, suggesting a strong effect. However, the sample is skewed by the presene of *aninit* and *arere*. Unlike *bipo* and *bihain*, these can only occur as spatial prepositions. To determine how preposition type determines the presence of *long*, the next sample is concerned only with *bipo* and *bihain*. Running the same tests, the phi-coefficient for the association in this sample is 0.615, suggesting a somewhat weaker, but still strong effect, which is still highly significant at $p < 0.001$. All of this seems to lead to the conclusion that a) spatial prepositions obligatorily occur with *long* and b) temporal prepositions optionally occur with *long*.

144

Having established that there is a link between the type of preposition and the presence of *long*, the next logical question to ask is why that should be the case. Is it due to a fundamental functional difference, or are structural features the decisive factor? The answer might lie in the different possible prepositional complements that spatial and temporal prepositions can take. While spatial prepositions will usually take a noun phrase as a complement, temporal prepositions can take either a noun phrase or a clause as a complement.

Form	NP	CL
bihain (+*long*)	64	1
bihain (-*long*)	62	123
bipo (+*long*)	41	6
bipo (-*long*)	35	168
aninit (+*long*)	250	-
aninit (-*long*)	-	
arere (+*long*)	250	-
arere (-*long*)		-

Table 6.8: Occurrences of different prepositional objects of *bihain, bipo, arere* and *aninit* with and without *long*

In conceptual terms, a noun phrase often stands for an entity - animate or not - while a clause can indicate an event. Coding the same sample as before for type of complement clause, a pattern emerges as seen in 6.8. While *aninit* and *arere* only occur with NPs (as predicted), both *bihain* and *bipo* occur with clauses as prepositional complements as well. Furthermore, when they occur with clauses, they seem to favour dropping *long*. If the presence of *long* is determined by whether or not the prepositional complement is a noun phrase or a clause, this indicates that rather than conveying a spatial notion, *long* is used as an entity/object marker instead. The (constructed) examples below illustrate the difference:

(19) Em i   aninit   long   **haus**.
      3SG PM PREP-L PREP house
      It is underneath the house.

(20) Em i   bipo   long   **haus**.
      3SG PM PREP-L PREP house
      It is in front of the house.

(21) Em i   kam bipo   (long) **yu**.
3SG PM come PREP-T (PREP) 2SG

He came before you.

(22) Em i   kam bipo**mi i   lusim**.
3SG PM come PREP-T 1SG PM   leave

He came before I left.

In the sample for *aninit*, *arere*, *bipo* and *bihain*, the phi-coefficient for the association between the prepositional complement type (NP or CL) and the presence of *long* is, at 0.584, weaker than the one between prepositional type and the presence of *long* for both samples above. However, it still indicates a strong effect, which is significant at $p < 0.001$. We might be able to draw the following conclusion: the presence of *long* within the complex prepositions surveyed is correlated with a) the type of preposition (spatial or temporal) and b) the type of prepositional complement, with the caveat that the two factors might be strongly influencing each other: the association between prepositional type and prepositional complement comes in at a phi-coefficient of 0.729.

What, then, does this tell us about *wantaim* and how it came to shed *long*? The link between *long* and NPs versus clauses seems to not be informative in this regard: *wantaim* as a preposition has NPs as complements exclusively, so the syntactic form of the complement seems not to be a factor. As implied above, though, *wantaim*'s weaker link to the spatial dimension might, given that all spatial prepositions require *long*, whereas temporal ones do not.

An additional issue which has affected the development of other simple spatial prepositions might be competing serial verb constructions. For Tok Pisin, Verhaar (1990, 119) reports the following, among others[15]:

(23) Yu wantaim ol soldia bilong yu i   mas **wokabaut**
2SG PREP-C PL soldier PREP-P 2PL PM must walk-around
raunim   dispela taun wanpela taim long   olgeta de.
around-VB DT   town NUM   time PREP DT   day.

You and your soldiers must march around this town once every day.

(24) Lea wantaim ol pikinini bilong   en i   mas **wokabaut**
Lea PREP-C PL children PREP-P 3SG PM must walk-around
bihainim tupela.
follow   PRN.

Lea and her children had to follow the two [girls].

---

[15]For both examples 21 and 22, glossing mine and translation supplied by original source.

In these examples, the second part of the serial verb construction serves a spatial or directive function which, in English, would have been fulfilled by a preposition. A similar competing serial verb construction that would function as comitative has, to my knowledge, not been attested. In addition, Sebba (1987, 214) notes that "most if not all of the serialising languages [...] have prepositions corresponding to *with* and *of* or *for*." This indicates that there may actually be a common tendency for the comitative to be expressed by prepositions rather than serial verbs.

This would mean that in the overall development of *wantaim*, *long* actually served as kind of a catalyst. Its presence allowed a temporal adverbial to be grammaticalized into a comitative preposition. Once that function was achieved, factors such as the productivity of *wan*+X and the frequency of the construction as a whole has led to increasing independence from *long* and the spatial context it implies, effectively eliminating the need for *long* at some point.

As with *ol*, the point of emergence for *wantaim* is significant in terms of language external factors as well. There are no attestations for comitative or instrumental *wantaim* during the jargon and stabilization stages. Being cut off from its main lexifier - and therefore the formal supply - led to internal developments in Tok Pisin, such as the reanalysis of a temporal adverbial as a comitative preposition. The input change from L1 English to L2 English would have made a direct borrowing of a comitative preposition from English more unlikely.

If we assume that the development of comitative *wantaim* has occurred as detailed above, one question still remains: why did Tok Pisin not follow one of the typologically more common grammaticalization paths? These would include, according to Luraghi (2014, 109f.), space or location as a source domain as well as the 'comrade' or 'people' domain in addition to serialized verbs with meaning such as 'follow' or 'take' and from adverbs with the meaning of 'together'.

Heine and Kuteva (2002, 329) list the following three concepts as sources for the comitative: COMRADE, FOLLOW and TAKE. For the first, they note that "the notion 'comrade' stands for a number of role relations, including 'companion', 'friend', 'neighbour', 'relative', though they suspect that this may be an "areally confined phenomenon" that has not been attested for non-European languages (Heine and Kuteva 2002, 91f.). In any case, Tok Pisin lacks a common single lexical item for 'neighbour' and 'relative'[16], while 'friend'

---

[16]They can be expressed by circumlocution as in *man long lain bilong en* 'man of the same family' or *narapela susa na brata* 'another sister and brother', the latter indicating that

and 'companion' are likely to be expressed by *pren*. At 4,693 occurrences, *pren* is a fairly common word in the Z'graggen corpus. However, it mostly occurs in syntactic environments that seem unfavourable to reanalysis as the comitative.

A productive syntactic environment for reanalysis would put *pren* in a position where it would be ambiguous as either a noun or a preposition. A sentence such as **Em i tokim pren man bilong narapela ples* 'He talked with a male friend from another village', if it occurred, would provide such an opportunity. This is, however, not the case, and *pren* does not occur in such ambiguous environments. In 2,521 occurrences alone, it is part of a possessive construction with *bilong*, as in *Na em askim pren bilong em* 'And he asked his friend'. In sentences of this type, *pren* could not be reanalysed as comitative, given that *bilong* needs a preceding noun. In turn, in almost all of the occurences of *tok+pren*, the sentence is akin to *Na em tok, Pren, yu laik ronowe* 'And he said, Friend, you want to run away', which do not allow for a comitative reanalysis of *pren*, either.

For the concept of FOLLOW leading to the comitative, Heine and Kuteva (2002, 140), describe it as "an instance of a process whereby process verbs, on account of some salient semantic property, give rise to grammatical markers expressing case relations", citing examples from Ainu, Mandarin Chinese and contemporary Chinese. The salient semantic property of FOLLOW verbs would be that successfully following someone ensures that one occupies roughly the same space (with minimal distance) at the same time. In Tok Pisin, 'follow' is expressed by *bihainim*, which, at 9193 occurrences, is the 116th most common word in the corpus. One of the main reasons for *bihainim* to not become grammaticalized as a comitative is likely to be its syntactic position. If sentences such as *Mama bihainim papa bilong pikinini bilong em go long hap* 'The mother followed the father of her children and went to the place', in which constructions such as N+*bihainim*+N occurred, were common, that would provide a potential environment for syntactic reanalysis of *bihainim* as the comitative.

However, that is not the case. *Bihainim* is only very rarely preceded by nouns: the most common are *meri* at 48 instances, *man* at 31 and *pikinini* at 16. Furthermore, *bihainim* is clearly marked as a verb by the transitive suffix *-im*. No *-im* verb has undergone grammaticalization in Tok Pisin. All nine TMA markers, for instance, derive either from nonverbal sources, intransitive verbs, or, in the case of *wok* and *pinis*, the unmarked version of the transitive verbs. *Bihainim*, however, needs the *-im* suffix to mark it as a verb, given

---

the person is not part of the immediate family.

that it is also very frequent as an adverb and a preposition. It is conceivable that such clear verbal marking provides an inhibiting factor for reanalysis into another functional word class.

The concept of TAKE leading to the comitative is thought by Heine and Kuteva (2002, 140) to be another instance of process verbs becoming grammatical markers. They note that "the exact conceptual nature of the present process is not yet entirely clear". In any case, Tok Pisin's verb for 'take' is *kisim*. At 65736 occurrences, it is the 25th most frequent word in the corpus and the third most common verb after *stap* and *tok*. Several of the inhibiting factors posited for *bihainim* could also have played a role in *kisim*'s failure go grammaticalize.

For one, like *bihanim*, *kisim* rarely occurs in N+*kisim*+N constructions: its three most common nominal 1L collocates are *meri* at 452 occurrences, *man* at 339 occurrences and *mama* at 245 occurrences. Similarly, *kisim*'s reanalysis could be inhibited by its clear marking as a (transitive) verb. In addition, one might expect that due to its semantic properties, *kisim* would be followed mostly by inanimate nouns rather than animate ones. As discussed above, we expect newly grammaticalized patterns to emerge with animate nouns first. However, this does not seem to be the case: two of the five most common 1R nominal collocates of *kisim* are, in fact, *pikinini* at 976 occurrences and *meri* at 790, which are certainly animate nouns (the other three are *paia*, *pis* and *kaikai*).

To summarize, I have shown a possible grammaticalization path for *wantaim (long)* from a temporal adverbial to the comitative preposition first and the instrumental preposition second. Due to a variety of factors, *wantaim* shed *long* during its grammaticalization process, in contrast to other complex prepositions which still carry it. The timing of the grammaticalization process once again played a role in denying the speakers of an emerging more complex variety of Tok Pisin the formal supply of English with, requiring speakers to draw from existing, internal constructions. Other grammaticalization paths for the comitative were disadvantaged due to different factors. The main process in the grammaticalization was semantic reanalysis rather than syntactic reanalysis: the combining concepts of spatial and temporal alignment led to a rare, but not unique grammaticalization of a temporal construction to the comitative preposition.

6.3 Adjectival suffix: *-pela*

One of the characteristic features of Tok Pisin and its sister languages, *-pela* is mainly used to form adjectives. There is, however, some merit to debating whether *-pela* is an inflectional suffix or a derivational one. Mühlhäusler (1985, 93f.) treats it as part of inflectional morphology along with the transitive suffix *-im*. Verhaar (1995, 12), on the other hand, claims that "Tok Pisin has no inflectional affixes, it only has derivational affixes - and, once again, all of them are suffixes". Mühlhäusler (2003, 14) describes the function of inflectional morphology as being threefold: signalling word class membership; signalling secondary semantic information like tense, aspect or number; and signalling grammatical relations such as case. The question is relevant in so far as *-pela* would, should it be an inflectional affix, be the only one in Tok Pisin.

According to Verhaar (1995, 12), "the suffix *-pela* is used to form a limited numbers of modifiers to nouns [...], some pronouns [...], and a variety of numerals [...]." In Mühlhäusler and Wurm (1985, 93), the authors note that "[i]n stabilised Tok Pisin, as recorded in the years after 1900, its occurrence appears to be restricted" to either being "a marker of monosyllabic attribute adjectives" or "a marker of plurality with the first and second person pronoun mipela *we* (excl.) and yupela *you* (pl.)". However, Mühlhäusler later notes that "lexical derivation can shift morphological affixes from one grammatical category to another" (Wurm and Mühlhäusler 1985, 336) and that "the freedom with which Tok Pisin speakers can shift words from one word class to another makes these affixes rather unreliable indicators of word class" (Mühlhäusler 2003, 15). He cites the following examples as indicative of this fact:[17]

(25) Long**pela** bilong   dispela bris   i   fiti yat.
      Length   PREP-P DT   bridge PM fifty yard
      The length of this bridge is 50 yards.

(26) Em i   singaut bik**pela**.
      3SG PM shout   big.
      He shouts loudly.

(27) Yu   no   ken askim dis**pela** askim.
      2SG NEG can ask   DT   question
      You should not ask this question.

(28) Kainkain ples i   gat nara**pela** kolim bilong   dispela samting.
      All-sorts place PM have DT   name PREP-P DT   thing
      Every other place has a different term for this thing.

---

[17]Emphasis and glossing mine, translation from source.

As is obvious from these examples, *-pela* occurs not only, as described in Mühlhäusler and Wurm (1985, 93), as an adjectival marker and a plurality marker in a narrow set of personal pronouns, but also in items that function like nominalised adjectives, adverbs and determiners.

Aside from adjectives which "indicat[e] nationality, language and religious affiliation", Mühlhäusler (1985, 355) cites the following adjectives as not taking *-pela*:

Adjective	Translation	Adjective	Translation
belhat	hot tempered	*nogut*	bad
bruk	broken	*pas*	stuck
daun	low	*pret*	afraid
giaman	false	*siut*	right (side)
hait	hidden	*slek*	loose
hambak	vain, proud	*tambu*	forbidden
kais	left (side)	*tan*	done (of food)
kela	bald	*taranggu*	unfortunate
klia	clear(ed)	*tru*	genuine, real
kros	angry	*tulait*	bright
malomalo	soft	*yarpas*	deaf
marit	married	*yaupas*	deaf
nating	empty, useless, worthless	-	-

Table 6.9: Adjectives not taking the *-pela* suffix

The data in the Z'graggen corpus does not fully match this claim: while most, namely *belhat, giaman, hait, hambak, kais, kela, kros, malomalo, marit, nating, pas, pret, siut, taranggu* and *yarpas* indeed never take *-pela*, the others do (albeit very rarely):

Adjective	Occurrences with -*pela*	Occurrences without *-pela*
tru	130	25400
daun	78	23789
klia	14	3927
bruk	12	n/a
slek	6	n/a
nogut	5	11545
tambu	3	3707

Table 6.10: Number of occurrences of adjectives with *-pela*

Still, there does not seem to be any pattern as to the exceptions as far as I can discern. Furthermore, these adjectives occur a lot more frequently without *-pela* than they do with it. The variation does not seem to be lexically determined either. As Mühlhäusler notes, most of these adjectives follow the modified noun rather than precede it. This would seem to indicate that the presence of the *-pela* suffix is determined by syntactic position rather than a link in the semantic content of these adjectives. Neither does there seem to be a link in their etymological origin, given that some, like *tru* or *marit* seem to be direct borrowings from English, while others, like *nating* are semantically close borrowings from English adjectives and yet another group are former English verbs such as *hait*. Still others, like *yarpas*, are taken from various substrates.

A partial explanation for *-pela*'s presence or absence is be found in its diachronic development. In regards to the origin of *-pela*, Crowley writes:

> All three dialects of Melanesian Pidgin have a suffix that derives ultimately from the English noun 'fellow', which is reflected as *-fala* in both Bislama and Solomons Pijin. In Tok Pisin it is represented orthographically as *-pela* [...].
> The precursor to the suffix *-fala* was already in place in Early Beach-la-Mar. There is no evidence in written sources from the 1840s to the 1860s to suggest that there was anything but a form *felo*, which was a synonym of the noun *man* meaning 'person, man'. If there was a difference between these two forms, it was that *felo* is only ever attested as occuring when there was a preceding adjectival premodifier, whereas *man* is only found in written sources when the noun is unmodified, or when the premodifier was itself a noun. (Crowley 1990, 275)

There is, in fact, little doubt that the form *-pela* and its equivalents in Bislama and Solomon Islands Pijin derive from English *fellow*. Crowley's observation that *felo* was more likely to occur after an adjectival premodifier further strengthens the argument that *fellow* was subjected to syntactic reanalysis. Consider the early TP interpretation of an English sentence such as *He is a good fellow*. Likely, a speaker of Pacific Pidgin English would have translated it to something like *Em i gut pela*. Since the pidgin never acquired the indefinite article - which would indicate that the subject is part of a subset of good fellows - the meaning can easily be parsed as 'he is good'. In other words, the noun phrase is reanalysed as an adjective in predicate position.

So far, this would indicate that the origin of *-pela* lies in pure syntactic reanalysis. Certainly, there is no semantic element to *fellow* that would make it a good candidate for an adjectival marker. However, there might also have been categorial supply from a substrate language, namely Tolai. Crowley (1990, 284) expands on Faraclas' observation in this regard as follows:

> Faraclas (1988: 127-34) offers a tantalizing structural parallel between Tok Pisin preposed attributives and a similar construction in Tolai, which was one of the major early substrate languages. In Tolai preposed attributives are linked to a following noun by means of the particle *na*, while postposed attributives directly follow the noun, as in:

(29) *a   pal   pua*
      art house white
      'the white house'

(30) *a   pua   na pal*
      art white link house
      'the white house'

Crowley continues, however, by noting that "there is no widely distributed option for these kinds of attributives to occur prenominally in the substrate languages for Bislama, and there is no morpheme paralleling the *na* of Tolai". He therefore concludes that the fact that *-fala* still evolved to have the same status in Bislama as *-pela* did in Tok Pisin is indicative of the fact that substratum influence was not as central to the development as Faraclas has claimed.

As we have seen so far, *-pela* occurs both in attributive as well as predicative adjectives (see also Verhaar (1995, 13) and Hall (1943a, 20)). Mühlhäusler (1985, 93), however, claims that "[i]n stabilised Tok Pisin, as recorded in the years after 1900, its occurrence appears to be restricted to two functions", namely that of a marker of monosyllabic attributive adjectives only on the one hand and as a marker of plurality with the first and second person pronoun (*mipela* and *yupela*) on the other. Hall (1943a, 20) likewise notes that "[p]olysyllabic adjectives almost never take the suffix". While this constraint on polysyllabic adjectives seems to hold true in the Z'graggen corpus (not a single polysyllabic adjective occurs with *-pela*), *-pela* clearly occurs with adjectives in predicate positions (see examples 31 and 32 below):

(31) Pinga bilong   em i **longpela longpela** [...]
      Finger PREP-P 3SG PM long   long   [...]

His finger was **very long** [...]. (ZC 00569)

(32) Yes tupela i   **yangpela** i   pikinini yet.
      Yes PRN   PM young   PM children still

      Yes, both were **young**, they were still children. (ZC 00497)

The most common pronouns and adjectives occuring with *-pela* certainly tend to be monosyllabic, as the following table shows:

Item	Occurrences	Item	Occurrences
dispela (this)	154256	*traipela*	3071
tupela (two)	90636	*nupela*	2857
mipela (I)	51305	*tripela*	2522
wanpela (one)	47183	*arapela*	2448
bikpela (big)	21125	*sotpela*	1398
yupela (you)	20753	*waitpela*	1051
narapela (another)	13306	*retpela*	993
gutpela (good)	11383	*strongpela*	990
mitupela (we, exclusive)	11220	*olpela*	972
sampela (some)	11007	*naispela*	955
yutupela (the two of you)	7990	*ispela*[18]	857
yangpela (young)	4600	*foapela*	750
longpela (long)	3915	*blakpela*	659

Table 6.11: Most common items ending in *-pela*

The fact that *-pela* does not occur with all adjectives, but mainly monosyllabic ones, complicates an argument for categorial demand. If categorial demand was present and the driving issue behind the development of *-pela* into an adjectival marker, it must have been limited to a certain subset of adjectives for some reason. Furthermore, there is the issue of some monosyllabic adjectives occurring both with and without *-pela* in both adjectival and predicative position, such as colours like *blak(pela)* or *ret(pela)*:

(33) Skin bilong   ol   i   **blak** [...]
      Skin PREP-P 3PL PM black [...]

      Their skin was black [...]. (ZC 02149)

---

[18]Likely to be a frequent misspelling of *dispela*

154

(34) Dispela **blak** koki i kam long ples na i mekim
DT   black parrot PM come PREP-D place CONJ PM do
wanem?
what

This black parrot came to the place and then did what? (ZC 03943)

(35) Oh yu gat **blakpela** skin [...]
Oh 2SG have black   skin [...]

Oh you have black skin [...]. (ZC 03530)

(36) [...] kala bilong em i **blakpela**.
[...] colour PREP-P 3SG PM black

[...] its colour was black. (ZC 03245)

However, there seems to be a tendency for *blakpela* and *retpela* to occur in attributive position more frequently than in predicative position. The same holds true for *ret* and *blak*, which seem the preferred choice in the predicative position. *Ret*, for instance, only occurs in attributive position with *pen* 14 times and with *pis* once, out of a total of 771 occurrences. *Blak* occurs in attributive position one time each with *dokta*, *hat* and *Pater*, twice with *misin* and *wesan*, three times with *graun*, *pen* and *mazik*, four times with *si*, once with *kala*, 15 times with *koki* and 17 times with *skin*[19].

This hints towards the fact that the above theory of syntactic reanalysis holds merit: it is in predicative position that sentences such as example 36 above would have had *fellow* follow the adjective. It is also in predicative position that polysyllabic adjectives tend to occur with *-pela* more than in attributive position. In any case, it is highly plausible that *-pela*, as one of the characteristic elements of the Melanesian languages, emerged not through some process that is unique to Tok Pisin or its sister languages, but through the same syntactic reanalysis that happens in other languages as well, which is an observation that will repeat itself in the analysis to follow below.

6.4 Transitive suffix: *-im*

The suffix *-im* is used in Tok Pisin to form transitive verbs, i.e. verbs accepting one or more objects. There are several characteristics that *-im* shares with *-pela*: they are both features that appear in all three Melanesian languages, and they are both likely to be the result of syntactic reanalysis that has resulted in

---

[19]Note, though, that *blak skin* seems to be a compound noun.

opaque suffixes that probably are part of the derivational morphology of Tok Pisin rather than of an inflectional morphology. Furthermore, similar to *-pela*, according to Smith (2002b, 59), "the transitivizing marker ***im*** is one of the most chracteristic features of Melanesian Pidgin English, and its use has been recorded from the earliest Pidgings of the Pacific". About its' origin, he writes that it is "no doubt [...] the English object pronoun 'him'", noting that "this is sometimes said to be reanalyzed according to Melanesian substrate patterns (e.g. Keesing 1988), although the suffix was present in NSWPE considerably earlier". The use of [-im] as a transitive marker most certainly dates back at least to the Jargon Stage of Tok Pisin, as Mühlhäusler (1985, 93) notes:

> *Him*, found in a number of positions in Jargon English, has become phonologically differentiated, *em* being used as third person singular pronoun and *-im* as a suffix to verbs having an object. The exceptions to the latter convention, i. e. verbs such as *gat* and *kaikai*, are also established by around 1920. It appears that the distinction between verbs which are compulsorily marked by *-im* and others which can use either *long* or *-im* to mark transitivity also dates to this phase, though a more detailed investigation is needed to confirm this.

In the case of *-im*, however, this has not been entirely unmotivated syntactic reanalysis, given that, as Mufwene (1986, 141) points out, the "[u]se of a suffix to distinguish transitive and/or causative verbs from intransitive/non-causative verbs is one of the morphosyntactic features shared by most Melanesian languages". One example would be Tolai, which allows speakers to transform morphologically unmarked intransitive verbs into transitive verbs by, among other strategies, attaching the suffixes *-e* and *-a* (Mosel 1980, 42). Thus, the reanalysis of *him* as a transitive suffix has had substratum reinforcement in terms of categorial supply.

Mihalic (1957, 24) gives the following general rules for transitive verbs in Tok Pisin:

1. Any verb ending in the suffix *im* is transitive and has a direct object without any prepositions being required [...]
2. Any transitive verb not ending in the suffix *im* requires a preposition before its object [...]
3. The following are exceptions which though not having the suffix *im* can take a direct object without need of a preposition: *drink* 'to drink', *gat* 'to have', *kaikai* 'to eat', *pilai* 'to play', *save*, 'to know', *tekewe* 'to take away', *telimaut* 'to divulge'.
4. N.B. The verbs *drink, kaikai* and *tekewe* also have the usual transitive form ending in *-im*.

5. Compound Transitive Verbs:

6. Some transitive verbs when compounded with the adverbial suffixes *ap*, *aut*, *ewe* make use of the objective suffix *im* (or its equivalent: *am*) twice. Once it is infixed between the stem and the adverbial suffix; then it appears again at the end of the resultant compound [...]

7. The majority of transitive verbs, however, when compounded with adverbial suffixes *ap*, *aut*, and *ewe* do not add *im* twice.

It is the last point that is of special interest to me here: in some instances of forming complex verbs such as *liftimapim* (to lift up), *-im* is present twice. The first *-im* then constitutes what Verhaar (1995, 23) calls "a 'fossilized' *-im* [...] without difference in meaning". Its functional motivation is already served by the second *-im*. Examples 37 and 38 show, respectively, instances of the use of *liftimapim* and *pulimapim* from Jon Z'graggen's recordings. Example 39 shows an instance of *pulapim*, in which the first *-im* has been deleted:

(37) Na   ol **liftimapim** em putim long   bet nau.
     CONJ 3PL lift-up   3SG put   PREP bed now.

     And they lift him up and put him on the bed now. (ZC 01945)

(38) Em save **pulimapim** long   bilum.
     3SG HAB fill   PREP bag.

     He fills the bag. (ZC 02701)

(39) Orait   na   Garong i   **pulapim** dispela liklik hap   long
     Alright CONJ Garong PM fill   DT   small piece with
     mambu [...].
     bamboo [...].

     Allright and Garong filled this small piece with bamboo [...]. (ZC 00752)

Tok Pisin seems to be remarkably resilient to deleting one of the *-im* elements, with forms such as *liftimapim, tekimautim* and *holimpasim* being preferred up to this day for some forms, as 6.3 below shows. This raises two questions: first, which *-im* is functionless, if any? The first one, because *-im* is usually a suffix? The second one, because *-im* is part of the verbal part of the compound? One possibility is the question as to which of the two *-im*s is deleted more frequently. However, there is no clear pattern as to whether both *-im* elements are deleted and if not, which one is retained, as 6.4 shows. The second question would be what kind of compounds they are: V+ADJ or V+V. For *aut* and *ap*, for instance, both *autim* and *apim* occur as independent verbs:
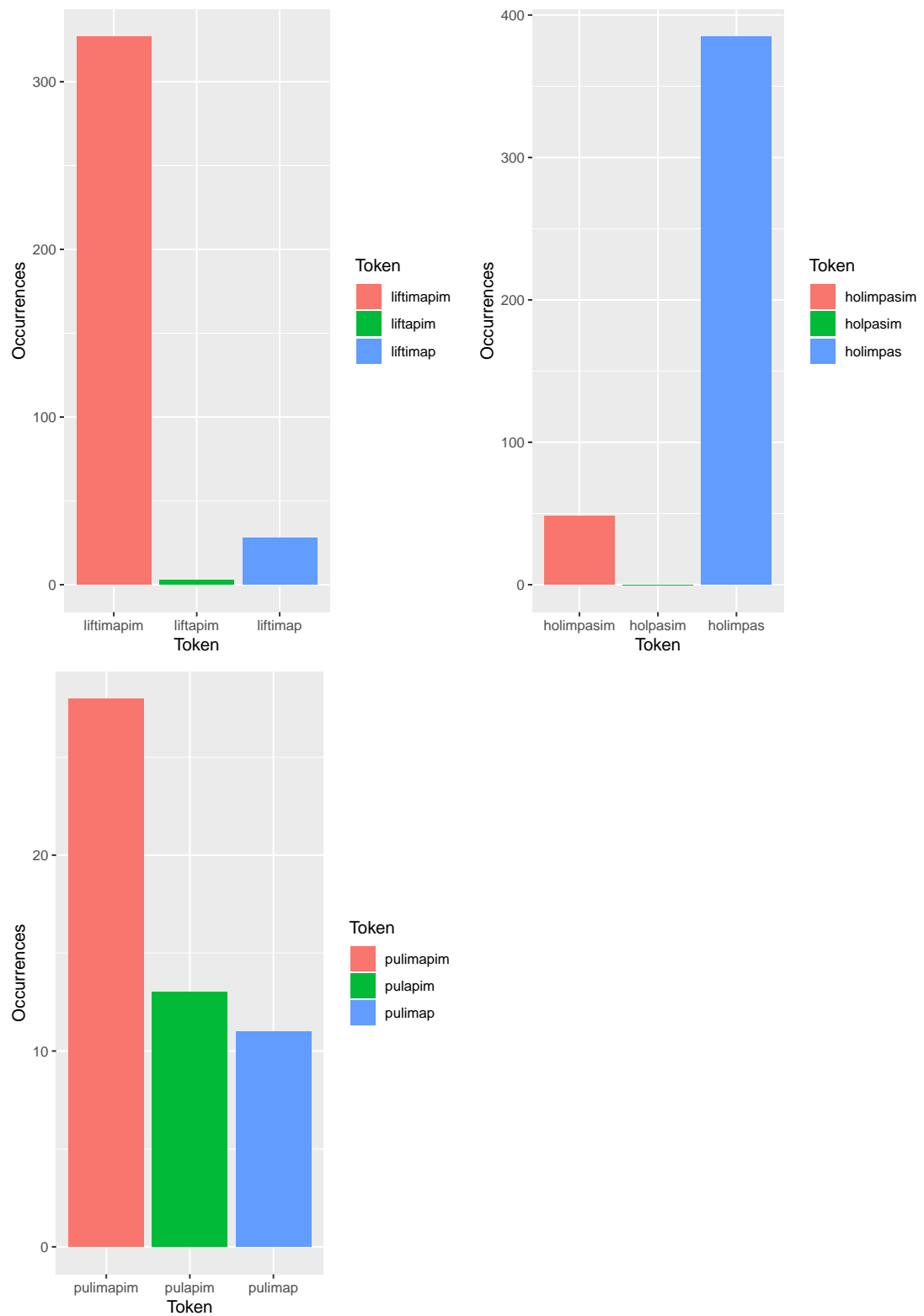
Figure 6.3: Occurrences of the various forms of three double *-im* verbs

(40) Orait   em nau God **i** **autim**   ol   long   dispela Paradis
      Alright, 3SG now God PM throw-out 3PL PREP-D DT   Paradise
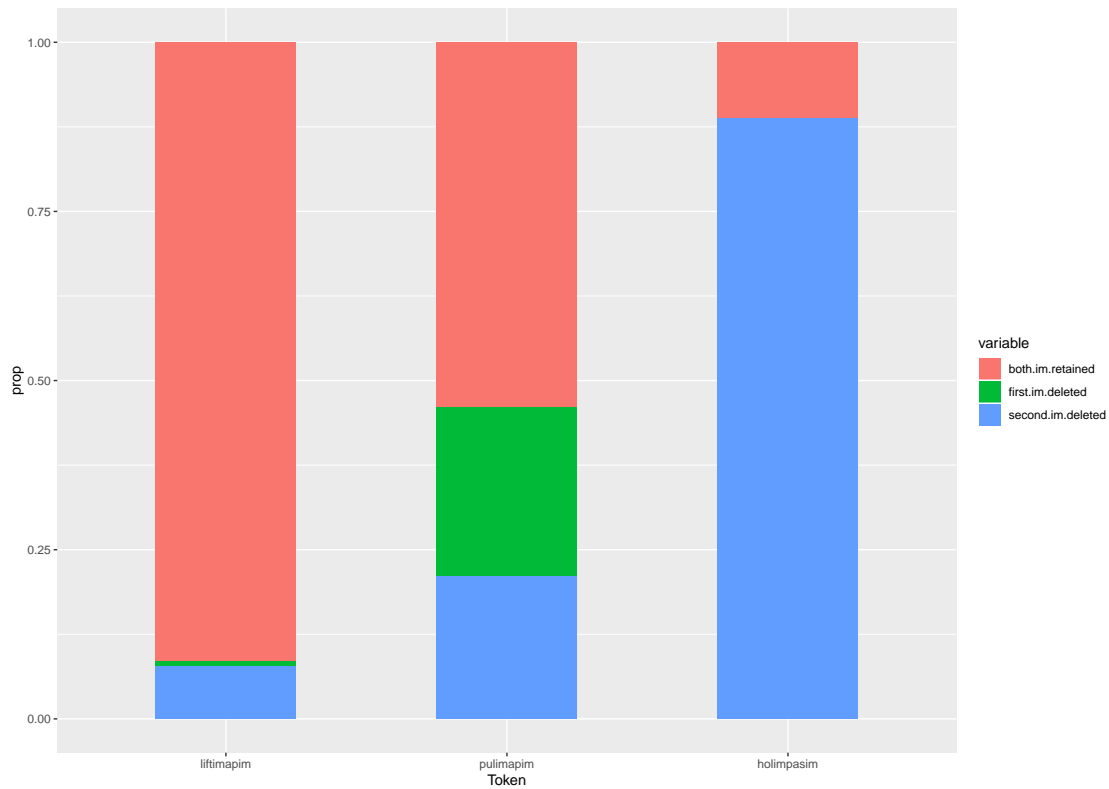
Figure 6.4: Deletion of *-im* in double *-im* verbs

na	ol	i	kamap	autsait,	Adam	tupela	Eva	ol	i	lusim
CONJ	3PL	PM	come-up	outside	Adam	DT	Eva	3PL	PM	leave

dispela Paradis. (ZC 04072)
DT   Paradise

Alright, then God threw them out of Paradise and they came outside, both Adam and Eve left this Paradise.

Assuming that these are double-verb compounds, this would indicate that *-im* is part of the lexeme, i.e. that it is either a derivational affix or is not transparently recognized as an affix at all. The fact that it is frequently retained even in middle position speaks in favour of such an analysis: if it were parsed as an affix, it would be unnecessary in this position. In any case, it provides us with a second instance of syntactic reanalysis - with possible substratum reinforcement - as the foundational process of introducing a new grammatical element into the emerging structure of Tok Pisin. It is of note that the only two affixes in the language resulted from this process.

6.5 Predicate marker: *i*

No survey of innovations among Tok Pisin would be complete without examining the predicate marker *i*, i.e. the marker preceding, in Tok Pisin, the finite verb of the clause and its auxiliary verbs. As Hall (1943a, 21) remarks, "this particle normally precedes the center of the predicate [...] and is proclitic to the first word thereof". He calls it a "third-person predicate marker", indicating that it is absent for the first and second person pronouns *mi, mipela, yu, yupela* and *yumi.* This may have been the case in earlier stages of the language; however, while still a tendency, in the Z'graggen data, the absence of *i* is no longer categorical for the type of personal pronoun, as table 6.12 shows[20].

Item	with *i*	without *i*	proportion
mi (1st SG)	3849	110402	0.035
yu (2nd SG)	3048	87948	0.035
em (3rd SG)	153479	496323	0.309
mipela (1st PL)	13269	51305	0.259
mitupela (1st PL excl)	3408	11220	0.304
yumi (1st PL incl)	1444	12088	0.119
yupela (2nd PL)	1996	7990	0.250
ol (3rd PL)	11383	28674	0.397

Table 6.12: Occurrences of pronouns with and without *i*

As Mühlhäusler (1990, 235) notes, "there remain a number of descriptive uncertainties when it comes to formulating the grammar of *i* in present day Tok Pisin", quoting Smeall (1973, 1) in saying that it "stands out, in the sea of polysemy which is Tok Pisin, as an element to which no functional status has been assigned with any success". There is little doubt, however, that two of its functions - whether they be the central, core or original function or not - are marking the predicate and functioning as a copula, as the examples below show. Examples 41 and 42 show pronouns plus *i* in copula constructions, whereas 43 and 44 show standard *i* with third person subjects:

(41) I   no   longpela stori. **Em i** liklik.
      PM NEG long   story. 3SG PM   small.

   It's not a long story. It is short. (ZC 02658)

---

[20]Note that since the corpus is not POS-tagged, the numbers in this table include all occurrences of the pronouns, including in an object position, where *i* would not occur.

(42) Bikos **ol i** liklik liklik na ol i strongpela na **ol i**
Because 3PL PM small small CONJ 3PL PM strong CONJ
karim i go antap.
3PL PM carry PM go on-top
Because they were very small and they were strong and they carried
upwards. (ZC 01007)

(43) No, ol i no go long ples, **ol i** go long dispela
NEG 3PL PM NEG go PREP-D place 3PL PM go PREP-D DT
raun wara [...]
round water [...]
No, they don't go to the place, they go to this lake [...] (ZC 00773)

(44) Orait **em i** go bek tokim ol man long ples Samsam [...]
Alright 3SG PM go back talk PL man PREP-L place Samsam [...]
Alright, he went back and talked to men at Samsam [...] (ZC 00997)

In regard to the origin of *i*, Hall (1966, 83) writes:

> The greater independence of the predicate in many pidgins and creoles is also
> shown by the fact that it is often set apart from what goes before, by some spe-
> cial syntactic marker. South Seas Pidgin English, in general, has a "predicate
> marker" /i-/, which is normally used when the subject (if there be one) is not of
> the first or second person: for example, Neo- Melanesian /mastɛr ĭ-si ŋawtcm
> mi/ "the master called me"; Neo-Solomonic /hem i-tekim loŋ sŏr/ "he took it
> to the shore". The form /i-/ occurs in no other function in the language and
> therefore cannot be considered a pronoun, although it unquestionably comes
> from English *he.* Its use reflects a merger of the substandard English habit of
> recapitulating a subject by means of a pronoun — as in John, he's an idiot or
> my mother, she always spanks me — and the Melanesian-Micronesian feature
> of morphologically distinct pronouns that recapitulate subjects and introduce
> predicates, as in Marshallese /ládrik e-gérabal/ "the boy, he works".

Mühlhäusler (1990, 240) points out that one of the few facts linguists who
have investigated the predicate marker agree on is that its origin precedes the
"crystallization of Tok Pisin as a separate language". He points out that any
attempt to look for its origin in simple substrate influence or borrowing from
languages such as Tolai would be fruitless. Similarly, Mosel (1980, 120ff.) re-
marks that while the origin of *i* is "obviously" from Melanesian substratum
influence, the fact that it is found in "ancient Bichelamar" and its descendants
rules out a Tolai origin. While the ultimate origin lies as far back as that, Mosel
further points out that is grammaticalization into the functions it serves today
happened later, in individual settings and with individual outcomes such as in

Tok Pisin.

Mühlhäusler (1990, 247) agrees, noting that before 1885, around which the split between Tok Pisin and its sister languages began to be established, and "for a considerable time thereafter", there was high variability of predicate marking, even though contact with Tolai and related languages increased. Since no uniform function and syntactic environment for *i* existed before the split, the predicate marker evolved along similar lines in all three languages, but is not identical in the present day. This, according to Mühlhäusler, "suggests that the emergence of predicate marking is unlikely to have been determined by a universal development hierarchy". In general, however, *i* works similarly in both Bislama (Crowley 2004, 110) and Pijin (Huebner and Horoi 1979, 12):

> In Bislama the form *i* and less commonly *oli* often appears between a subject and a following predicate. We therefore find examples such as the following (with the predicate markers presented in bold type in each case):
>
> Hem | **i** singsing.
> '(S)he is singing.'
> Roro **oli** save lukluk.
> 'Roro can see.'
>
> The i between the subject and verb in the above sentences is a predicate marker and marks the beginning of the verb phrase.
> It occurs optionally when the subject is second or third person (i.e. *yu*, *yufala*, etc.) (i.e., *hem*, *olketa*, etc.) or first person plural (i.e., *mifala*, etc.) There apparently is no difference in meaning between the (a) and (b) varieties of the following sentences.
>
> > 2a. Hem i ranawe.
> > 2b. Hem ranawe.
> > 'He ran away'.
> > 3a. Jon an Rut i save.
> > 3b. Jon an Rut save.
> > 'John and Ruth know.'

As hinted at above, there is considerable variance between the presence or absence of *i*. The following two tables illustrate this variance by comparing its presence after pronouns at the start of the sentence:

V/Prn	Mi	Yu	Em	Yumi	Mipela	Yupela	Ol
go (go)	18	22	5726	24	220	347	2716

stap (stay, be)	28	59	2475	14	88	93	1141
kam (come)	10	15	1821	2	35	126	1229
tok (talk)	0	0	457	0	3	0	776
kisim (get)	0	0	204	0	1	0	176
save (know)	1	0	263	2	37	5	439
kaikai (eat)	0	0	90	0	2	1	156
kamap (arrive, begin)	0	0	455	1	20	20	213
lukim (see)	0	0	277	0	10	5	241
mekim (make)	0	0	326	0	3	8	289

Table 6.13: Pronouns with -*i* (at the beginning of sentences)

V/Prn	*Mi*	*Yu*	*Em*	*Yumi*	*Mipela*	*Yupela*	*Ol*
go	593	2007	397	77	41	95	26
stap	241	357	253	12	14	37	35
kam	167	719	577	3	10	76	88
tok	218	155	6480	3	74	27	1562
kisim	187	595	2061	14	70	144	850
save	219	1496	819	28	318	76	637
kaikai	79	178	637	7	53	79	453
kamap	58	54	523	4	33	26	108
lukim	283	504	1120	3	43	107	335
mekim	73	314	793	11	27	84	517

Table 6.14: Pronouns without -*i* (at the beginning of sentences)

Note that *mi* does not trigger *i* very frequently, while *em* and *ol* do, (see Verhaar 1995, 71). In general, Verhaar's observation that among the subject pronouns, *i* is triggered by third person singular *em*, third person plural *ol*, first person plural exclusive *mipela* and second person plural holds true for the speakers represented in the Z'graggen corpus, as can be seen from tables 6.13 and 6.14 above. They show the ten most common verbs in the corpus and their occurrences with each pronoun with and without *i* at the beginning of sentences, respectively[21].

In general, *mi, yu, yumi* and *mipela* occur very infrequently with the

---

[21] The survey was limited to the beginning of sentences by searching for the capitalized version of the pronouns in the Z'graggen corpus, given that a case-insensitive search produced a very high number of false positives caused by constructions with *bilong* in sentences such as *Father bilong mi tok...* 'My father said (...)'

predicate marker. Exceptions are mostly limited to *i go*, *i stap* and *i kam*. Interestingly enough, the numbers for *em* run counter to expectations: while for *go, stap* and *kam*, the number of occurrences with *i* is significantly higher, the opposite is true for *tok, kisim, save, lukim, kaikai* and *mekim*. The numbers for *kamap* are relatively balanced.

There are various factors, including syntactic position and phonological environment, that influence the presence and absence of *i* in present-day Tok Pisin. What is of interest to me here, however, is not the variety at present, but the origin of the form. Much the same as *-im* and *-pela*, the predicate marker *i* is the result of syntactic reanalysis as described by Hall above. The fact that it is triggered more often in contexts including the third person is reflective of that origin; it occurs in the same syntactic position where the recapitulated subject in substandard English, as Hall refers to it, would have occurred, as well as in the same synactic position where the recapitulating pronouns of Melanesian languages occur.

The last three items surveyed have three things in common. First, they all derive from an English etymon, respectively, *fellow, he* and *him*. Secondly, they were all subjected to syntactic reanalysis, which formed the beginning point of grammaticalization. This means that they were not grammaticalized into a function that is related to their inherent semantics and concepts, but into functions which are related more closely to their syntactic position and syntactic environment. While not entirely random - the words occurred in said positions for a reason in the input - their semantic bleaching might have been more successful than if their semantics had played a larger role in the process. Thirdly, they are all features which emerged early in the origin and development of Tok Pisin or earlier pidgins. Thus, their development was influenced to a much higher degree by the original input than it is the case for other phenomena below.

6.6 Connectives: *na, tasol* and *o*

According to Haspelmath (2004, 5), there are three[22] semantic types of coordination, namely conjunction (also known as conjunctive coordination), disjunction (or disjunctive coordination) and adversative coordination. The main difference among them is, of course, not primarily linguistic in nature, but conceptual. Linguistically, they all link noun phrases, predicates or clauses.

---

[22]To be precise, he posits that there are at least three, noting that "sometimes an additional type "causal coordination" is distinguished".

Conceptually, connectives describe the nature of the relationship between these entities, states or events. Conjunction allows for both entities, states or events to co-exist, while disjunction limits the choice to one of the options. Adversative coordination does not quite fit in with the other two. It allows for both entities/states/events to co-exist, but a conflict arises on a different conceptual level. If one interlocutor were to ask the other, for instance, whether they had gone to the store to get groceries, the answer could be *Yes, but they were closed.* As such, the contrast for the adversative conjunction arises through new information, whereas the contrast for the disjunctive is preset. These diverse concepts can be reflected in the various grammaticalization paths that the conjunctional, adversative and disjunctive connectives can arise from, as we will see below.

Haspelmath further distinguishes between monosyndetic and bisyndetic coordination, which involve one single coordinating item or two coordinating items. A further typological difference is established between languages which place the coordinator after the first coordinand ([A co] [B]), before the second coordinand ([A] [co B]), after the second coordinand ([A] [B co]) or before the first coordinand ([co A] [B]). In Tok Pisin, all three semantic types of coordination exhibit monosyndetic coordination, with the coordinator occurring between the coordinands, as examples 45 through 47 show:

(45) Yu trikim mi **na** yu go.
     2SG trick   1SG CONJ 2SG go
     You tricked me and you left. (ZC 00039)

(46) Yu ken lukim mi **tasol** yu no   inap tok.
     2SG can see   1SG CONJ 2SG NEG ABL talk
     You can see me but you cannot talk. (ZC 02265)

(47) Yu kilim mi **o**   yu laik kaikai.
     2SG hit   1SG CONJ want food
     You hit me or you want food. (ZC 00721)

In addition, all of the coordinators consist of a single item. Therefore, the distinction between monosyndetic and bisyndetic coordination as well as between the placement of the coordinands is of minor relevance to this analysis and will not be discussed below.

Another common distinction in the analysis of connectives is that between nominal and verbal coordination. One might further distinguish between verbal coordination and clausal coordination. However, following Haspelmath's (2004, 10) example, the notion of verbal coordination adopted in this study

encompasses both clauses and verb phrases, given that they often cannot be clearly distinguished and are marked by the same grammatical items in Tok Pisin (for instance, both coordinative conjunction of verb phrases and of clauses is achieved by *na*).

6.6.1 Coordinative conjunction: *na*

As mentioned above, *na* is the coordinative conjunction in Tok Pisin, whose main function is to link predicates and/or clauses without establishing a contrast, as can be seen in examples 48 and 49 below. In a secondary function, it also links noun phrases, as is evident in examples 50 and 51 below.

(48) Mangi   ia   kaikai planti **na**   em slip nogut tru
     Young-man FOC eat   plenty CONJ 3SG sleep bad   really
     long   nait.
     PREP-L night

     The young man ate a lot and slept really badly at night. (ZC 03062)

(49) **Na**   ol i   abrusim **na**   ol i   go.
     CONJ 3PL PM miss   CONJ 3PL PM go

     And they missed and they went. (ZC 01028)

(50) Na   taim em i   kam, papa **na**   mama i   gat bikpela
     CONJ when 3SG PM come father CONJ mother PM get big
     hamamas.
     happy

     And when it comes home, father and mother get very happy. (ZC 03623)

(51) Orait   dispela Manub **na**   Kulbob, taim tupela i   kros.
     Alright DT   Manub CONJ Kulbob when PRN   PM angry

     Alright, both Manub and Kulbob, when both were angry. (ZC 04276)

Once again, Tok Pisin employs a form of a grammatical item that diverges from the typical form among many of the other English-lexified creoles in the APiCS. As is evident from table 6.15, most of the other contact languages find the source of their coordinative conjunction in English *and*.

Language	Coordinative conjunction	Source
Early Sranan	*en*	English *and*

Sranan	*en*	English *and*
Saramaccan	*hen*	English *and*
Nengee	*neen/da*	???
Creolese	*an*	English *and*
Trinidad Endlish Creole	*an*	English *and*
Vincentian Creole	*an*	English *and*
Jamaican	*an*	English *and*
Belizean	*n*	English *and*
San Andreas Creole English	*an*	English *and*
Nicaraguan Creole English	*an*	English *and*
Bahamian Creole	*and*	English *and*
Gullah	*and*	English *and*
African American English	*and*	English *and*
Krio	*ɛn*	English *and*
Ghanaian Pidgin English	*ɛn(d)/ɛn dɛn*	English *and (then)*
Nigerian Pidgin	*ànd*	English *and*
Cameroon Pidgin English	*an*	English *and*
Pichi	*àn/we*	???
Chinese Pidgin English	*and*	English *and*
Singlish	*and*	English *and*
Tok Pisin	*na*	???
Bislama	juxtaposition/*mo*	English *more*
Norf'k	*en*/verb chain	English *and*
Kriol	*en*	English *and*
Hawai'i Creole	*æn*	English *and*

Table 6.15: Cumulative verbal conjunction in the 26 English-lexified languages in the APiCS data

It is interesting to note here that Bislama also diverges from the common pattern. However, it does not employ the same coordinative conjunction as Tok Pisin. Instead, it uses either the juxtaposition of sentences or a coordinative

conjunction *mo*, which most likely derives from English *more*[23]:

(52) Tom i   go long   maket **mo** hem i   pem tu   manggo.
     Tom PM go PREP-D market CONJ 3SG PM buy NUM mango

     Tom went to the market and bought two mangos. (Crowley 2004, 173)

(53) Wan ston i   rol long   hil **mo** i   brekem haos blong
     DT stone PM roll PREP-D hill CONJ PM break   house PREP-P
     mifala.
     1PL-EXCL-MULT

     A rock rolled down the hill and smashed our house. (Crowley 2004,
     173)

Solomon Islands Pijin, on the other hand, falls squarely in line with most of
the other contact languages and makes use of either *and* or an adapted form,
as is evident from the two examples below[24]:

(54) [...] mi ting iu   mas go long   hospital long   Auki kwik
     [...] 1SG think 2SG must go PREP-D hospital PREP-P Auki quick
     taem **and**   tekem samfala meresen.
     time CONJ take some   medicine

     [...] I think that you should go to the Auki Medical Centre [and] get
     some medicine. (Regional Assistance Mission to Solomon Islands 2011,
     17)

(55) Olketa stap long   antap maonten **an** weitim nao [...]
     3PL   be   PREP-L on-top mountain CONJ wait   now

     They are up on the mountain waiting [...] (Regional Assistance Mission
     to Solomon Islands 2011, 11)

I have been suggesting that *na* does not derive from English *and*, despite re-
peated claims to this effect in non-linguistic works. Muysken (2008, 196), in
an overview of Tok Pisin conjunctions, question words and relativisers, lists
the origin of *na* with a question mark. Most other overviews and grammars
do not go into detail with regard to *na*'s origin at all, either accepting it as
an adaptation of *and* or not deeming its origin an issue worth pursuing. Since
na functions much like English *and*, this might not be all that surprising: why
would it have a more complex origin, when such a seemingly simple answer

---

[23]For each of the following examples, translation from the original source, glossing mine.
[24]For each of the following examples, translation from the original source, glossing mine.

suffices? Even if the form does not derive from English, its functionality, after all, still might. In other words, we might have another case in which categorial supply was present and so was formal supply. At some point, so must have been categorial demand. The question, then, is if formal demand was present and when it was.

After all, it would be theoretically possible that *na* (/na/) has its origin in English *and*. We have to consider that the strong version of English æ*nd* can be reduced to as many weak forms as /ənd/, /ən/, /nd/, /n/, /m/ and /ŋ/ (Jones and Roach 2009, 22). In addition, Jones notes that *and* is a "weak form word", whose strong form is mostly only used for emphasis or citation and that the most frequent reduced form is /ən/. Add to that the fact that in its function as a nominal conjunction, it would have been likely to occur between two content words in a sentence and therefore even less likely to receive stress as a verbal conjunction. Thus, from either /ən/ or /n/, for instance, it is possible to reconstruct a path of phonological change that results in Tok Pisin's *na*.

The most likely path, given both the frequency of the /ən/ weak form in the English input and final-consonant deletion as a phonological process[25], would have had to undergo either metathesis and subsequent vowel strengthening to /nə/ and later to /na/ or further reduction to /n/ and subsequent reintroduction of a reduced vowel. Both of these paths seem unlikely, given that the function of the word would not have changed and it would remain in the same unstressed position it did before.

To add to that, Hall (1943a, 29) notes that "[t]he missionary text (Tok bilong Baibel) occasionally uses *an(d)*, which is clearly an Anglicism, and not 'true Pidgin' at all; B writes: 'Your conjunction *and* - I never heard it, and don't believe in it'".

On its own, given the regional differences between the varieties of Tok Pisin, such a statement would of course hardly discredit the notion that *na* might stem from *and*. In addition, since *na* and *and* are sufficiently different phonologically, the person who provided the statement above may have found no link between these two words even if they were of one and the same origin. In combination with the evidence above, however, a picture begins to emerge that hints at a different origin story for *na*.

Consider also how Bislama does not use an *and*-derived form for its coordinating conjunction. Of the three Neo-Melanesian dialects, only Solomon Islands Pijin does. This is another indication that *and* was, at the very least,

---

[25]A process that certainly does occur in Tok Pisin as well, for example in *graun* < E *ground* or *han* < E *hand*.

not adopted before the split of the three Neo-Melanesian dialects. Bislama and Tok Pisin then developed a coordinative conjunction based on material existent within the languages - *na* in Tok Pisin and *mo* in Bislama.

The origin of *na*

Rather than deriving it from *and*, I suggest that Tok Pisin *na* as a coordinating conjunction very likely originated as a reduced form of *nau*. First of all, the path of phonological change from /nau/ to /na/ is much shorter and less complex, requiring only the monophothongization of /au/ to /a/. Note, however, that vowel monophthongization is not necessarily a prevalent feature in Tok Pisin. Diphthongs are preserved in a multitude of words, both including those of English origin such as *bia, bihain, taim, nait, boi, dai, dais, daun, dia, drai, foa, gia, graun, hait, haus, kaun, krai, laik, nait, paia, yau, yia*, etc. and those adopted from German or substrates such as *rausim, buai, kaikai, kaiser, kiau*. In others, such as *de* < E *day, hul* < E *hole, we* < E *where , nem* < E *name, nil* < E *nail, pen* < E *pain, pepa* < E *paper, pes* < E *face, page, ples* < E *place*, plet < E *plate, pre* < E *pray, pret* < E *afraid, ren* < E *rain, sop* < E *soap, sos* < E *sauce, spet* < E *spade, tel* < E *tail*, the diphthong is monophthongized.

There seems to be a tendency to preserve /aɪ/ and to reduce /eɪ/ to /e/, but none for /au/ to be monophthongized to /a/. The reduction from /nau/ to /na/, should it have happened, would therefore have to be attributed to the frequency, position and function of /nau/ rather than to a general tendency of phonological adaptation.

A similar process was noted by Mihalic (1957, 87) who writes, in reference to the distinction between *moa* and *mo*, that "[his] feeling with this word is that at the end of an utterance it is *moa*" and that "[w]hen a word follows 'more', it is *mo*", ultimately deciding to record it as *moa* "for safety's sake". It seems likely that a similar phonological reduction process has occurred with *nau*. If not in prominent and therefore stressed position at the beginning or end of a sentence, it was reduced to *na*. If the data from the Z'graggen corpus is any indication, *na* is very likely to have occurred in these positions. After the predicate marker *i* and the third person singular pronoun *em*, it is the single most frequent token in the corpus, occurring no less than 450,721 times. Of those 450,721 occurrences, about 113,000 (or, roughly, 25.2%) occur sentence-initially, and therefore in a position more likely to receive some measure of sentence stress. The remaining 74.8%, about 337,000 occurrences, are positioned within sentences. Naturally, this number includes *na* both as a verbal and a nominal conjunction, while the former is unlikely to contain a

function of nominal conjunction, explaining some of the stark discrepancy in numbers. As function words, both sentence-internal verbal conjunctions and nominal conjunctions are far less likely to receive sentence stress (Cutler and Foss 1977, 1), contributing to the likelihood of /naʊ/ being reduced to /na/. So far, it therefore seems far more likely for *na* to be a reduction of *nau* than to be an adapted version of *and*.

Furthermore as mentioned above, in the Z'graggen corpus, both forms - *nau* and *na* - still occur at the start of sentences within a narrative sequence, in the meaning of *and then*, as illustrated by the following passage:

(56)   **Nau** em i   tok, Sori pren, em yet   i   tingting. Em
        CONJ 3SG PM say   Sorry friend 3SG already PM think.   3SG
        yet   i   tingting na   tok, Oh sori   bai   mi   kisim dispela
        already PM think   CONJ say   Oh sorry FUT 1SG get   DT
        pikinini long   em. **Nau** em i   askim, Wanem taim tru   bai
        child   PREP 3SG. CONJ 3SG PM ask   what   time true FUT
        yu   redim   diwai bilong   wokim haus tambran bilong   pikinini
        2SG prepare tree   PREP-F craft   house ancestor PREP-P child
        bilong   mitupela?.   **Na**   em i   tok, Dispela taim long
        PREP-P 1PL-EXCL?. CONJ 3SG PM say DT   time PREP
        wanem taim tru, mi   no   save, bai   mi   ken redim   diwai.
        because time true 1SG NEG know FUT 1SG can prepare tree.
        Na   taim em i   go pinis,   ol   lain   brata   bilong   em i
        CONJ time 3SG PM go COMP PL family brother PREP-P 3SG PM
        kirap   ol   katim diwai. (ZC 00726)
        get-up 3PL cut   tree.
        And then he said, 'Sorry friend', he has already thought. He has already thought and said, 'Oh, sorry, I will get this child from him'. **And then** he asked, 'Really, when will you prepare the tree for building the house for the grandparent of our child?. **And then** he said, 'This time when, I don't know, I will be able to prepare the tree'. **And then**, when he was gone, his brothers in the family got up, they cut down the tree.

In addition, the original spelling in one of the texts in Mühlhäusler (2003, 61), uses *nao* for both functions, providing another clue of a common origin[26]

(57)   Buk   i   ko log   Smit tete   mi   kam apa lo   iu
        Letter PM go PREP-D Smith today 1SG come up   PREP 2SG
        log   gut Pel orait log   got Papa iu   mi   ol   tu   gat
        PREP good ADJ alright PREP get father 2SG 1SG 3PL also got

---

[26]Translation by Mühlhäusler, glossing mine.

peles **nao** was Matiu tede iu orait No Piko i
place CONJ wait Matthew today 2SG alright CONJ Piko PM
orait Iu tel im Sion **Nao** milaik tok iu log
alright 2SG tell TRA John CONJ 1SG-want say 2SG PREP
kamap Lomi tupel log Kewinag. [...] **Nao** i laik orait
come-up Lomi both PREP Kewinag. [...] CONJ PM want alright
likalik tede orait **Nao** mi laik Lapalap.
small today alright CONJ 1SG want clothes

This letter that went to Smith arrived today from you, (saying that) you are fine, that your father was fine, the two of us also have a place and we are waiting. Matthew, are you fine today? And Piko are you fine? Tell John and I want to tell you that (I and) Lomi, the two of us, arrived in Kevieng. [...] Now it is improving a little bit, today it is all right. Now I want a loincloth.

Mühlhäusler, in updating the text to modern spelling, transcribes the first two instances of *nao* as *na* and the third and fourth as *nau*. A further indication that *nau* and *na* are not only closely related, but are identical even for current speakers of Tok Pisin, is given by Smith (2002b, 186), who comments that "there is also some confusion over the identity of *na* and *nau*", noting that they are "not distinguished by some speakers". He goes on to say that "sequences of clauses are often punctuated by the us of both *na* and *nau*". This mixed usage is found in the Z'graggen corpus as well:

(58) Orait ol Australia, Amerika i kam, ol Misinari, **nau** yu
Alright PL Australia America PM come PL missionary CONJ you
Misinari long mi **na** i kam bek, **nau** mi man
missionary PREP-D 1SG CONJ PM come back CONJ 1SG man
gen **na** nau mi Independent, **nau** mi yet mi gat
again CONJ now 1SG independent CONJ 1SG already 1SG have
misin, mi yet mi gat gavman, em wok bilong God.
mission 1SG already 1SG have government 3SG work PREP-P god
(ZC 04009)

Alright the Australians, the Americans came, the missionaries, **and** you missionaries to me and I came back and I was a man again and I was independent now, and I already had a mission, I already had a government, it was the work of God.

The discourse-marking function of *na* is also recorded by Mihalic (1957, 90):

na

1) and, or. See *no* for "or", which is much more common.

Mi go na mi lukim. = I went and saw it.

Na wanem! = Of course!

Tupela boi na tripela boi i-mekim faipela. = Two boys and three boys are five boys.

2) so, and so

Em i-paitim mi na mi ronewe. = He hit me, so I ran away.

Skul i-pinis na mi go long ples. = School was over, and so I went home.

Note that in Mihalic's account, the first entry for *na* indicates its function as both a verbal and nominal conjunction, while the latter seems to have a causal nature: the speaker runs away because she is hit in the first sentence, and goes home because school was over in the second. This differs slightly from the second function Mihalic (1957, 93) records for *nau*, which he indicates to be "then - in narratives in the past tense", giving the following example[27]

(59) Ol   i-go,   i-go,   i-go,   **nau**   ol   i-kamap   long   Wewak
      3PL PM-go PM-go PM-go CONJ 3PL PM-come-up PREP-D Wewak
      They kept on going and going, then they reached Wewak.

Hall (1943a, 110) has separate entries for *na* and *nao* (which he spells *naw*). The entry for the former reads "conj 'and; then, so'" and notes that two of his five informants for spoken data distinguish it from *naw*, while the other three do not. For the meanings of *naw*, he records "the present time" as well as a secondary function as a "quasi conjunctive" with the meaning 'then, so, and'. Note that he specifically fronts the coordinative function in the entry for *na*, while he fronts the sequential function for *naw*.

Interestingly, *now* developed into another direction in Bislama, where it serves as a focus marker (Crowley 2004, 159ff.), which, in turn, is often shortened to *na*. This provides another indicator that *nao* and *na* in Tok Pisin have a common origin by proving the phonological reduction is viable.

All these indicators combined allow us to make a strong argument for an origin of *na* that lies with *nau* (and further back, with *now*) rather than with *and*. The suggested development would be from an adverbial *nau* with the meaning 'now', as directly borrowed from English to acquire an additional function of a sequential conjunction of sorts with the meaning 'and then'. Following this stage is an overlap between *na* and *nau*, where in certain positions,

---

[27]Translation his, glossing mine.

nau is reduced to *na*, resulting in two distinct grammatical items with shared function. Finally, *na* acquires the full function of a nominal and verbal conjunction. This development is displayed schematically below:

$$nau \Rightarrow \text{'now'}$$
$$\Downarrow$$
$$nau \Rightarrow \text{'now, and then'}$$
$$\Downarrow$$
$$na(u) \Rightarrow \text{'now, and then'}$$
$$\Downarrow$$
$$nau \Rightarrow \text{'now, and then' PLUS } na \Rightarrow \text{'and then'}$$
$$\Downarrow$$
$$nau \Rightarrow \text{'now, and then' PLUS } na \Rightarrow \text{'and then, and'}$$
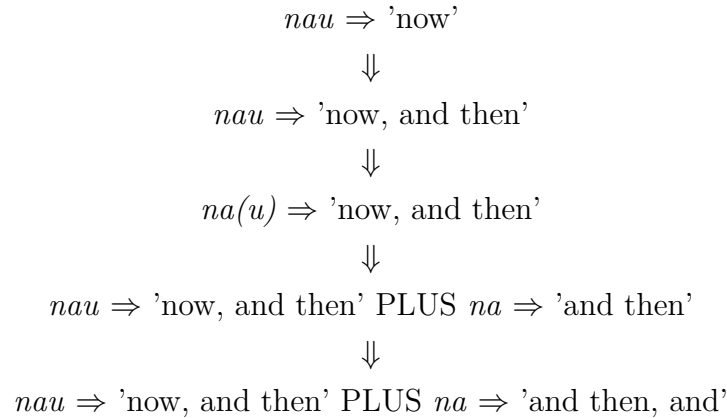
Naturally, *nau* and *na* are no longer functionally identical. While *nau* can only be used for verbal conjunction, *na* functions for both verbal and nominal conjunction, likely because it acquired the function after the two had split into two grammatically distinct items. *Nau* is much more closely linked to an event structure, i.e. a sequence of events or actions, not entities that are formalized as nouns.

> In English and other European languages, the coordinators 'and' and 'or' can link a diverse range of categories: noun phrases, verb phrases, clauses, adjective phrases, prepositional phrases, and others. The coordinator 'but' is mostly confined to clauses, but this seems to be for semantic reasons.
>
> But many languages have category-sensitive coordinating constructions [...]. In particular, about half of the world's languages show different conjunctive constructions for nominal and verbal/clausal conjunction [...]. (Haspelmath 2004, 9)

Said tendency is partly responsible for the fact that while *na* is the only possible verbal/clausal conjunction, nominal conjunction in Tok Pisin can be achieved by either *na* or *wantaim*, as illustrated by the examples below:

(60) Manub **wantaim** Kulbob, ol   i   kros   long   meri,   meri
      Manub CONJ   Kulbob 3PL PM angry PREP woman woman
      bilong   Manub.
      PREP-P Manub

      Manub and Kulbob, they were angry with the woman, Manub's wife.
      (ZC 00874)

(61) Meri **wantaim** pikinini meri bilong em.
Woman CONJ child woman PREP-P 3SG

The woman and her child. (ZC 03549)

Haspelmath (2013) notes that "quite a few creole languages show the differentiating pattern [...] with two different coordinators". He goes on to suggest that out of the two, it is generally the nominal conjunction marker that is an innovation based on or identical to the comitative preposition, while the verbal conjunction marker is older. So far, that picture matches with Tok Pisin. In regards to its origin, he continues saying that "[t]he use of a conjunction marker that also (and originally) means 'with' and that is restricted to nominal conjunction is a striking example of a substrate-influenced feature" citing Michaelis and Rosalie (2000) as an example. They develop the following model (what they call "la carte sémantique générale diachronique")[28]:

allatif humain

⇓

destinataire ⇐ bénéficiaire

⇓

successif ⇒ conjonctif ⇐ comitatif ⇒ instrumental ⇒ agent du passif
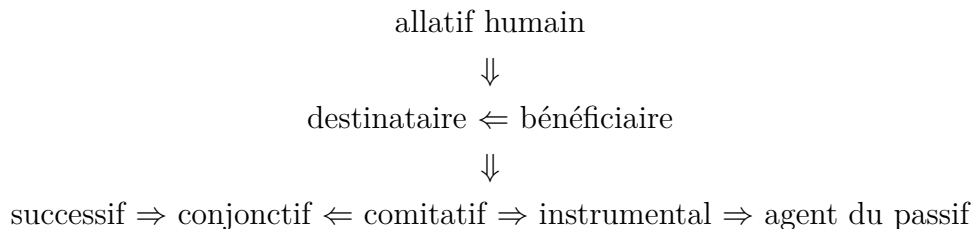
Haspelmath's own semantic map for conjunctions and related notions looks very similar (2004, 19):

```
                                  existence        manner
                                     |               |
V-conjunction — N-conjunction — comitative  —  instrumental
                       |                             |
                    'also'                         agent
                       |                             |
                    'even'                       comparison
```
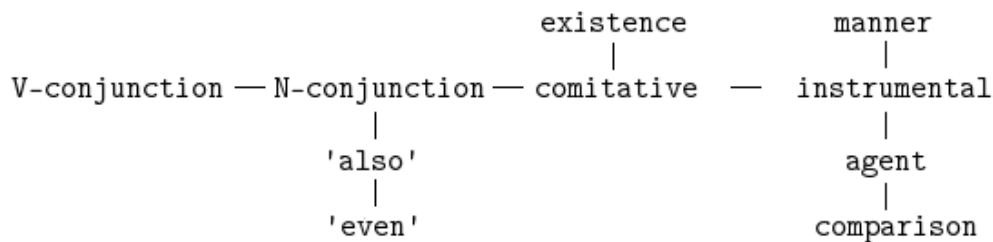
Figure 6.5: Semantic map for conjunctions

On this map, there is a path from conjunctive functions to the comitative. Accordingly, there is another factor that links *na* to the discourse-structuring function of *nau*, namely that of its distribution with *wantaim*, which, as mentioned above, can also serve as a coordinate conjunction, albeit only for nominal conjunction. This is hardly unusual, as Haspelmath points out:

---

[28]Figure adapted and simplified from Michaelis and Rosalie (2000, 90).

There are at least two different ways in which this formal identity can be understood. On the one hand, one can argue that the comitative/conjunctive markers in WITH-languages have just one single function, which happens to be rendered in two different ways in AND-languages like English that must differentiate between 'and' and 'with'. [...] On the other hand, one could argue that the comitative marker and the conjunctive marker are different synchronically, both semantically and syntactically, and that the identity of their shape is due to a very common semantic-syntactic change from comitative marker to conjunctive coordinator. Of course, it is quite possible (and actually very likely) that some WITH-languages are of the former type, while others are of the latter type. (Haspelmath 2004, 13f.)

Tok Pisin falls squarely into the category of languages in which the comitative marker and conjunctive marker are different synchronically, but since *wantaim* is a newer form, it has not replaced *na* in all instances of coordinating noun phrases yet. However, *wantaim* has an advantage in the competition between two forms: the comitative has no sequential nature and is therefore better suited to nominal conjunction. In both its comitative and instrumental function, *wantaim* necessarily occurs with a following noun. *Na*, on the other hand, is hampered by its close link to a sequential nature. Sequence implies a certain distance, whereas the comitative, as described above, implies equal time and space. Consider the semantic map by Haspelmath above: while both grammaticalization paths to the nominal conjunction are equally valid, the distance between the comitative "node" and the nominal conjunction "node" is not the same as the distance between the verbal conjunction "node" and the nominal conjunction "node". Intuitively, the link between verbal conjunction and nominal conjunction seems stronger to me conceptually, and the distance therefore smaller, than between the comitative and nominal conjunction. And indeed, in a sample of 500 instances of *na* as a functional conjunction, 480 occurrences could be classified as verbal conjunction, while a mere 20 were classified as nominal conjunction. On contrast, all 500 occurrences of *wantaim* as a conjunction in the equivalent sample connected noun phrases rather than predicates or clauses, as figure 6.6 shows. Note, however, that *na* is much more frequent than *wantaim* overall. Even if only 4% of its occurrences are nominal conjunctions, it would still be more frequent than *wantaim* in that function. For instance, in a sample of nominal conjunctions connecting two nominal phrases with proper nouns, *wantaim* occurred 882 times, while *na* occurred 2,727 times. Thus, while it confirms that *wantaim* has only progressed along the grammaticalization path from comitative to NP-AND or nominal conjunction on Haspelmath's map (and in the opposite direction to the instrumental),
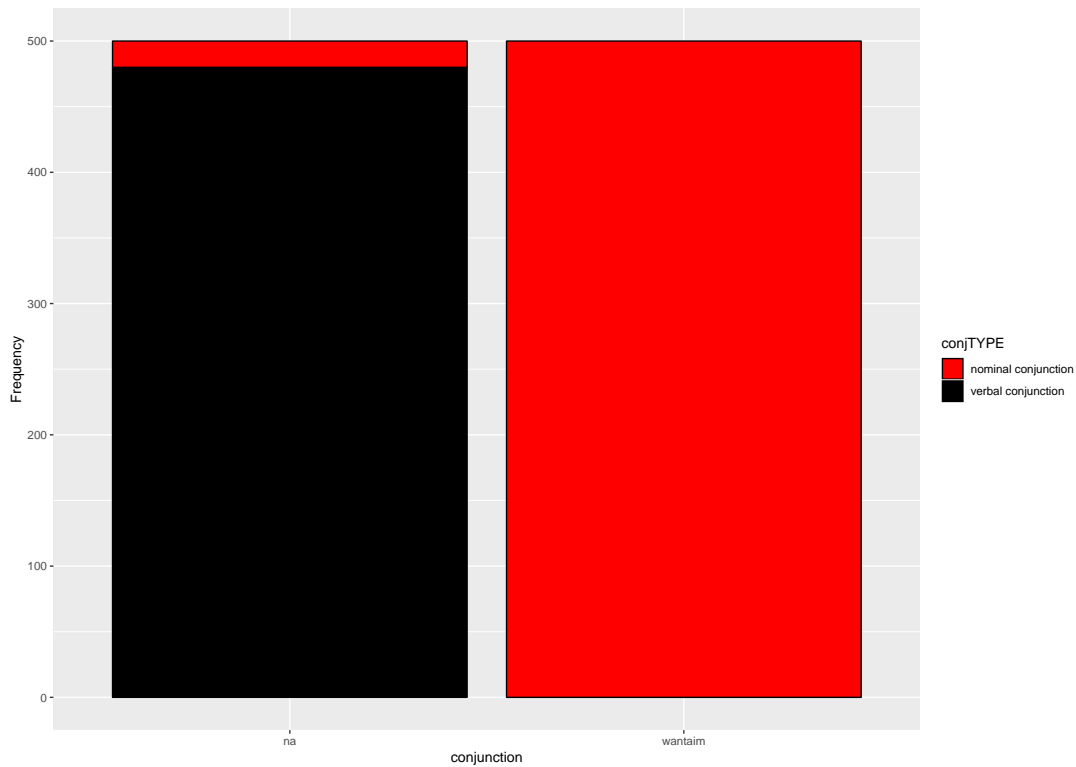
Figure 6.6: Occurrences of *wantaim* and *na* as nominal and verbal conjunctions

na has progressed in the same direction from V-conjunction to N-conjunction.

It is also noteworthy that there is hardly any difference between the nominal and pronominal types the two forms connect. The most common nominal and pronominal 1L collocates of *wantaim* are *meri, mama, pikinini, man, papa, mi, yu, mipela, susa* and *dok.* For *na*, they are *em, ol, tupela, mi, yu, man, meri, mipela* and *mama.* On the other side, the most common nominal and pronominal 1R collocates for *wantaim* are *ol, em, meri, mama, pikinini, man, papa, mi, yu* and *mipela*, while for *na* they are *em, ol, mi, yu, man, meri, mipela, mama, pikinini* and *papa.*

The grammaticalization path from comitative to "NP-AND", as they refer to it, is well attested in Heine and Kuteva (2002, 80ff.) and "appears to be well established" (ibidem, 82). The main difference in parsing NP coordination and the comitative seems to be the "ranking" of the NP constituents: whereas with the comitative, the conjoined NPs are separate entities, in their coordinated form, they are parsed as a single NP. As Stassen (2000, 26) puts it, "the grammaticalization of a comitative encoding pattern into a "coordination-like" construction prototypically involves the creation of a single constituent, in which both the 'with'-phrase and the non-comitative NP are included, and in which the two NPs gradually come to be regarded as being of equal structural rank". They also mention NP-AND markers as being one possible source for

S-AND markers (clause-conjoining markers), which leads them to speculate that "we may be dealing with a more general evolution COMITATIVE > NP-AND > S-AND". This seems to be the case with coordinating *wantaim* in Tok Pisin, which appears to have gone from comitative preposition to noun phrase conjoining marker (nominal conjunction) to clause conjoining marker (verbal conjunction)[29]. It remains to be seen, then, whether *wantaim* will expand its function to verbal conjunction as well. For now, it seems to be the sole domain of *na*, likely due to its link to sequential narrative structure.

This clear functional tendency also is one of the factors that speak in favour of an origin of *na* in the discourse marker *nau*, as opposed to a borrowing from substrates. Admittedly, there is a possible source in Tolai, which has a connective particle *na* as well. However, rather than serving as a coordinating conjunction linking noun or verb phrases, it instead links nouns and adjectives in attributive position, as in *a ko **na** gunan* 'a good village' or *a ngalangala **na** tung* 'big ravines'. (Mosel 1980, 104) It is also used in Tolai to form verbal nouns such as *tata na kankan* "from kankan to be angry [...] connected by the connective particle *na* with the preceding intransitive verb" (Mosel 1980, 95). Furthermore, it connects cardinal numbers and nouns in constructions such as *a laptikai **na** gai* 'six months'. None of these functions would explain the functional divide between *na* and *wantaim* as a discourse origin can, however. Still, it is possible that the substrate provided structural reinforcement for the form of *na* either through this construction or a near-homophone *ma*, which actually does serve as a coordinative conjunction, as shown the examples below (taken from Mosel 1980, 108, 122) show[30]

(62) ma   i   vana ma   i   vana ma   i   vana ma.
      CONJ 3SG go   CONJ 3SG go   CONJ 3SG go   CONJ
      And she go and she go and she go.

(63) Ma   i   ga ngala na pa.
      CONJ 3SG TA big   C taro
      And it became a big taro.

It is not inconceivable that its close phonological resemblance to *na* has benefited the grammaticalization of *na* by providing structural reinforcement, particularly given the fact that it seems functionally identical to Tok Pisin *na* as

---

[29]As a potentially amusing aside, the example Heine and Kuteva cite for the path from comitative preposition to S-AND is Swahili *na*, whose form is identical to Tok Pisin's *na* - except in Swahili, the *na* that ends up functioning as S-AND is the comitative preposition, whereas in Tok Pisin, *na* has its origin in the other path to the S-AND.

[30]Translation for both taken from source; glossing for first example by me; glossing for second example by source, here TA refers to tense/aspect marker, C to connective particle.

a verbal conjunction. Another possibility would be a common source for *na* in both languages, as it seems to serve a linking function in both.

Furthermore, there are a number of languages which also use *na* as a coordinating conjunction, among others Lakota (de Reuse 1982, 154), several Bantu languages including Luganda (Crabtree 1902, 58), and Swahili (Mohammed 2001, 249). However, since none of these languages can be considered substrates of Tok Pisin, they are of limited use in determining the origin of *na*. While the form does occur in several Oceanic languages, with several different functions between them, none of them employ it as a connective. In Fijian, for instance, it serves as a determiner (Maria Polinsky and Eric Potsdam ress, 16)[31]:

(64) na   cei e   a   savata na i   sulu?
      DET who 3SG PAST wash   DET NMLZ clothes

      Who washed the clothes?

(65) na   cava e   a   kania 'o   Jone
      DET what 3SG PAST eat   DET John

      What did John eat?

Finally, German also has an identical form *na* in colloquial use. It is an interjection that can function as a greeting in sentences such as *Na, wie gehts?* 'Hey, how are you?' or announcing imminent work in sentences such as *Na dann packen wirs mal an!* 'alright, let's get going then!'. This, however, is unlikely to have played a large role as well. First, it occurs only at the start of utterances, not in successive sentences as *nau* does. Secondly, it would be the only marker in Tok Pisin grammaticalized from German.

As was mentioned in the introduction to this chapter, there are certain conceptual differences between the adversative connectives and their conjunctive and disjunctive counterparts. These are summarized succinctly in Ramat and Mauri (2011, 657). They boil down to three main differences, the first of which is that "adversative connectives show a higher intra-linguistic variation than conjunctive and disjunctive connectives". Basically, speakers of a particular language have more options - that is, more grammatical items or patterns - to express an adversative relationship between two entities, states or events than they would have to express conjunction or disjunction. The example put forth by Ramat and Mauri is French, which "only shows *et* for conjunction and *ou (ou bien)* for disjunction, but a number of different connectives for contrast relations, e.g. *toutefois, mais, par contre, alors que, pourtant.*" Secondly,

---

[31] Glossing and translation taken from original; NMLZ stands for nominalizer.

they claim, "adversative connectives are more easily and quickly renewed than conjunctive and disjunctive ones, which instead seem more stable over time". Once again, they put forward Romance languages as an example, given that these have preserved both Latin *et* and *aut* in various forms, but none of the various Latin adversative connectives (*sed, tamen, at*). As an additional point, they pur forward Italian *però*, which was only grammaticalized in its current function during the 16th century. Their third point is that "adversative connectives are more easily borrowed than disjunctive and conjunctive ones", an argument that goes back to the implicational hierarchy put forth by Matras (1998, 301-305). As Ramat and Mauri summarize, this hierarchy predicts that "in bilingual contexts languages replacing combination markers also replace alternative (disjunctive) markers, and languages replacing alternative markers also replace contrast (adversative) markers". In other words, 'but' would be replaced before (or more quickly than) 'or', which in turn would be replaced before (or more quickly) than 'and'.

Matras attributed this tendency to a difference in "intensity with which the speaker is required to intervene with hearer-sided mental processing activities" (1998, 305) in the three different connective strategies. In short, adversative discourse relies on a higher contrast between the expectations of the interlocutors. Or, as Ramat and Mauri put it, "the more the relation implies a contrast, the more the speaker has to maintain assertive authority despite the denial of the addressee's expectations", which in language contact can be one causal factor in speakers adopting adversative conjunctions of the dominant language. Naturally, when connective strategies emerged in Tok Pisin, the dominant language was no longer English. Neither was it German, however, which means that speakers had to rely on strategies of their own devising to satisfy the need for renewal - or introduction, in the first stage - of an adversative conjunction.

6.6.2 Adversative conjunction: *tasol*

The second connective of interest to the present analysis is the adversative conjunction *tasol*. As with *na*, its origin at first seems immediately evident in English *that's all*, with certain phonological adaptations during its adoption into Tok Pisin. Unlike *na*, however, this origin holds up under closer scrutiny. Still, the grammaticalization path of *tasol* offers some valuable insights into grammaticalization and innovation in Tok Pisin, as I will show below.

To understand *tasol*'s emergence, it is first necessary to take a look at its

modern usage. *Tasol* serves multiple functions in modern Tok Pisin, which Inoue (2007, 233) summarizes as the grammaticalized expression *em tasol*, as well as functions as an exclusive particle, an emphatic particle and an adversative conjunction. An example for each function (in the above order) from the Z'graggen corpus follows below.

(66) Em i   go daun olsem   long hap bilong   nambis, em i   go
      3SG PM go down like-that to   place PREP-P beach   3SG PM go
      daun long   solwara. **Em tasol.** (ZC 00050)
      down PREP-L sea.   PRN   all.
      He went down like that to the beach place, he went down to the sea. That's all.

Inoue (2007, 233) notes that "*[e]m* in this expression is a demonstrative pronoun in Tok Pisin" and that "'that's all' is already grammaticalized [...] in Tok Pisin", frequently occuring "at the end of the discourse". Indeed, a large number of instances of *em tasol* in the Z'graggen corpus occur at the end of a transcription, with the narrating person using it almost as a formal construction to end the story that they were telling. It can therefore be referred to as a discourse marker, akin to the sequential discourse marker function of *nau* analysed above. Unlike *nau*, though, its function is not to introduce an additional or new sequence, but to signal that the narrative sequence has come to an end.

The next example shows *tasol* in the exclusive function:

(67) Long   nait **tasol** i   kamap.
      PREP-L night ADV PM appear.
      They appear only at night. (ZC 01500)

In this function, *tasol* occurs after the argument that it modifies, similar to how as a discourse marker, it appears after the sequence whose end it signals. The narrator is asserting that these particular flowers only appear/bloom at night, and not during the day.

An example of *tasol* in the emphatic function follows below:

(68) Na   em i   no   tumbuna   stori, em **nau tasol** mipela
      CONJ 3SG PM NEG grandparent story 3SG now   just
      lukim long ai.
      3PL   see   PREP-I eye
      And it is no story of our grandparents, we just now saw it with our eyes. (ZC 03521)

This usage of *tasol* emphasizes that it is only now that the event has happened, not at any other point in time. As such, it could be analysed as an overlap between the emphatic and exclusive function.

Next, an example of *tasol* in the function that is of primary interest to the present analysis, i.e. as an adversative conjunction:

(69) Mi   laik sutim balus **tasol** em i   abrisim na   i   kam
      1SG want shoot pigeon CONJ   3SG PM   pass-by CONJ PM   come
      pundaun long   garden bilong   yu.
      fall   PREP-D garden PREP-P 2SG

      I wanted to shoot a pidgeon, but it passed by and fell down into your garden. (ZC 00916)

As is evident from example 69 above, in its function as an adversative conjunction, *tasol* operates almost identically to its English equivalent *but*, including its syntactic position. It can appear both clause-initially as well as clause-medially and serves to establish a contrast between the two constituents it links. Constituents linked in this way are mostly clauses, though *tasol* can also establish a contrast between noun phrases.

Having presented the various functions of *tasol* in modern Tok Pisin, we shall first take a lot at whether *tasol* or structurally similar items occur in other English-lexified creoles within the APICs data in table 6.16.

Language	Adversative conjunction	Source
Early Sranan	*mara/nosu/tog/toku*[32]	???
Sranan	*ma/noso/toku*[33]	???
Saramaccan	*ma*[34]	???
Nengee	*ma/toku/ma toku*[35]	???
Creolese	*but*[36]	English *but*
Trinidad Endlish Creole	*but*[37]	English *but*
Vincentian Creole	*but*[38]	English *but*
Jamaican	*bot*[39]	English *but*

---

[32](van den Berg, Margot C. and Smith 2013, 11)
[33](van den Berg, Margot C. and Smith 2013, 11)
[34](McWhorter and Good 2012, 151)
[35](Goury and Migge 2003, 153f.)
[36]Devonish and Thompson (2010)
[37](Mühleisen 2013, 67)
[38](Prescod 2013, 77)
[39](Farquharson 2013, 88)

Belizean	*but*[40]	English *but*
San Andreas Creole English	*bot*[41]	English *but*
Nicaraguan Creole English	*bot*[42]	English *but*
Bahamian Creole	*but*[43]	English *but*
Gullah	*but*[44]	English *but*
African American English	*but*[45]	English *but*
Krio	*bɔt*[46]	English *but*
Ghanaian Pidgin English	*ba(t)*[47]	English *but*
Nigerian Pidgin	*bòt*[48]	English *but*
Cameroon Pidgin English	*bot*[49]	English *but*
Pichi	*pero*[50]	Spanish *pero*
Chinese Pidgin English	unmarked/juxtaposition[51]	-
Singlish	*but*[52]	English *but*
Tok Pisin	*tasol*[53]	English *that's all*
Bislama	*be*[54]	English *but*
Norf'k	???	???
Kriol	*bat/ani*[55]	English *but/only*
Hawai'i Creole	*ba(ʔ)*[56]	English *but*

Table 6.16: Adversative verbal conjunctions in the 26 English-lexified languages in the APiCS data

---

[40](Escure 2013, 98)
[41](Bartens 2013c, 111)
[42](Bartens 2013b, 124)
[43](Hackert 2013, 137)
[44](Klein 2013, 145)
[45](Green 2013, 154)
[46](Finney 2013, 165)
[47](Huber 2013, 174)
[48](Faraclas 2013, 183)
[49](Schröder 2013, 192)
[50](Yakpo 2013, 201)
[51](Matthews and Li 2013, 211)
[52]Lim and Ansaldo (2013)
[53](Geoff P. Smith and Jeff Siegel 2013, 220)
[54](Meyerhoff 2013, 229)
[55](Schultze-Berndt et al. 2013, 249)
[56](Velupillai 2013, 259)

As we can see, Tok Pisin once again deviates from the general pattern, which is to borrow *but* from English and adapt it phonologically. In this case, Bislama and Solomon Islands Pijin also show distinctly different adversative strategies from Tok Pisin. In Bislama, the equivalent conjunction is *be*, as examples 70 and 71 show. Like many other adversative conjunctions in English-lexified creoles, this form is likely adapted from English *but*.

(70) Tom i   go long   stoa **be**   Joseph i   go long   maket.
      Tom PM go PREP-D store CONJ Joseph PM go PREP-D market.

      Tom went to the shop but Joseph went to the market.[57] (Crowley 2004, 174)

(71) Hariken i   blu strong **be**   i   no   karemaot   haos
      Hurricane PM blow strong CONJ PM NEG knock-down house

      blong   mifala.
      PREP-P 3PL

      The cyclone blew hard but it didn't knock down our house. (Crowley 2004, 174)

In Solomon Islands Pijin, the adversative conjunction is *bat*[58]:

(72) Las naet brata   bliong   mi wande go fising **bat**   hem i
      Last night brother PREP-P 1SG want   go fishing CONJ 3SG PM

      hevi   ren.
      heavy rain

      Last night my brother wanted to go fishing, but there was heavy rain.
      (Huebner and Horoi 1979, 145)

(73) Mi   traehad fo   fosim haosben blong   mi fo
      1SG try-hard PREP-F force husband PREP-P 1SG PREP-F

      mitufala go long   sip **bat**   taem ia   hemi had tumas
      3PL   go PREP-D ship CONJ time FOC 3SG hard much

      fo   faendem rum   long   sip [...]
      PREP-F find   room PREP-L ship

      I tried hard to force my husband to have us go on the ship, but times
      were hard for finding room on the ship [...] (Holm 1994, 536)

As we can see, the function and position of the adversative conjunction remains the same across all three Neomelanesian dialects. The form, however, differs between Tok Pisin and the other two: while both Bislama and Solomon Islands Pijin adopt forms of English *but*, Tok Pisin does not.

---

[57]For both examples 70 and 71, translation from original source and glossing mine.
[58]For both examples 72 and 73, translation from original source and glossing mine.

Once more, we are confronted with the question of why Tok Pisin takes the road less travelled in this regard. One possible explanation, as usual, would be substrate influence: there might be structural supply from one of the languages involved. If so, the question arises as to whether that structural supply affected *tasol* in its emphatic, exclusive or conjunctional function. As for the exclusive function - from which the emphatic usually arises, which will be further discussed below - there was likely no substrate influence. In Tolai, the concept of *only* is expressed through the form *(u)ka*, which, like *tasol* in Tok Pisin, is postposed (Franklin et al. 1974, 119)[59]:

<u>ka</u>

(1) verb transitive 'to scrape, scratch' (2) abbreviation of <u>kan</u> (adverb 'perhaps'; conjunction 'last')

(3) 'only'

<u>Ia ka</u> It only

<u>U kaka u tar pait ia.</u> You only, you did it.

<u>I bolo ka.</u> He is only passing by.

<u>U kaka?</u> Are you by yourself?

So far, it seems like there was at least functional substrate influence, if not structural supply. Still, *ka* does not serve as an adversative conjunction. In fact, I have been unable to find information on how this function is expressed in Tolai. I therefore cannot conclude with certainty that there is no parallel structure or item in the Tolai language that could have influenced the development of *tasol*.

What about possible substrate influence from the Oceanic substrates? As Sankoff (1999, 11) notes, emphasizing and focus is often accomplished through postposed affixes or particles in various Oceanic languages. However, I posit that there is a simpler explanation for said development on a grammaticalization path that is well attested and that explains all four different functions of *tasol* in Tok Pisin as they were detailed above. To establish it, we will first have to look at the diachronic order in which these functions emerged.

For *tasol*, Mihalic's (1957, 147) entry already records both 'only, alone, just' in sentences such as *Em tasol; nau mekim gen* 'that's it, now do it again' and 'but, however' in sentences like *Mi laik go, tasol mi no ken* 'I want to go, but I cannot' as possible meanings. So does Hall (1943a, 121) 14 years prior, whose entry records both adverbial 'only' and "quasi-conjunctival" 'but, however'. As a discourse marker, *tasol* goes back to at least 1913, as recorded

---

[59]See also 1980, 102

by a letter printed in the *Koloniale Rundschau* and reproduced in Mühlhäusler et. al (2013, 50).

It seems prudent to assume that a discourse-marking *tasol* is the earliest function, given that it could have been adopted from English without any other necessary changes to its function. The next step would then have led to the expansion of function to the exclusive function of 'only' by simple semantic reanalysis: if the exclusive element is what is "all", then, vice versa, there is only the exclusive element. Consider example 68 above, which readily allows for both readings. The emphatic use follows naturally from its exclusive function: by narrowing the scope from all possible points in time to one single point - *nau tasol* - the speaker emphasizes what the scope is being limited to. It is only in the last step, then, that *tasol* would have gone from particle to conjunction. The suggested development is summarized below:

$$tasol \Rightarrow \text{'that's all'}$$
$$\Downarrow$$
$$tasol \Rightarrow \text{'that's all, only (exclusive)'}$$
$$\Downarrow$$
$$tasol \Rightarrow \text{'that's all, only (exclusive), only (emphathic)'}$$
$$\Downarrow$$
$$tasol \Rightarrow \text{'that's all, only (exclusive), only (emphathic), but'}$$
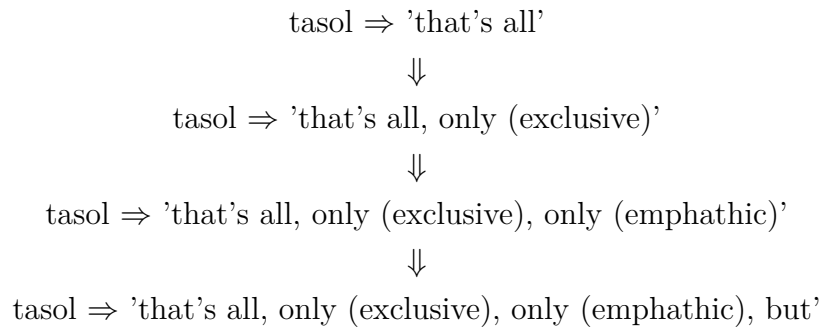
This development is in line with Inoue's (2007, 239) work. She shows the ambiguity between the exclusive, emphatic and adversative readings of *tasol* which made the reanalysis of each step possible. Ambiguity is, after all, a central requirement for reanalysis - if an item or pattern cannot be parsed in the new function, it would never be reanalysed. Inoue's proposal is therefore worth quoting in full:

> I propose that the exclusive particle became reanalyzable as an adversative conjunction via a stage as an emphatic particle. This hypothesis is based on what the synchronic pragmatic ambiguity shows us about the diachronic path of semantic expansion. Such semantic link among an exclusive particle and an emphatic particle and an adversative conjunction is not only found in Tok Pisin. There is a cross-linguistic link between these three functions. English *only* and French *seulement* also have these three functions and typically have the similar syntactic preferences. Below is an example of English only. In exclusive function, the scope of *only* tends to be NP. In emphatic function, the scope of *only* is VP. In adversative function, the scope of only must be a clause.

(12) I wrote only my family name. (Exclusive particle)

Scope of *only* ⇒ my family name (NP)

(14) It will only make her mad. (Emphatic particle)

Scope of *only* ⇒ make her mad (VP)

(15) I'd love to go, only I'm too busy. (Adversative conjunction)

Scope of *only* ⇒ I'm too busy (Clause)

In other words, what we are seeing during the diachronic development of *tasol* through the steps of the exclusive particle to the emphatic particle to the adversative conjunction is a gradual expansion of scope. This is especially interesting given Lehmann's (2015) grammaticalization parameter of reduced syntactic scope, which would have us expect the opposite development. Brinton and Traugott (2005, 138) already note, however that "discourse markers do not exhibit the scope reduction" and "the notion of scope reduction has been challenged in grammaticalization generally". In this cas the scope expansion affects not only the grammaticalization of discourse markers themselves, but possibly even the grammaticalization of further items that arise from these discourse markers. The same is, incidentally, true for the grammaticalization of *na(u)* as described above. Like *tasol*, it goes from being a discourse marker - in this case, a sequential discourse marker - to expanding its scope to verb phrases and functioning as a verbal conjunction. Similarly, *na* expands its scope from verb phrases to noun phrases.

If we assume that scope reduction is a feature of grammaticalization despite growing counter-evidence, then there must be a unique or at least unusual factor present in the grammaticalization processes of Tok Pisin. Among these factors are the syntactic complexity and any existent structures within an emerging language. For instance, we might ask whether scope reduction is likely in a language which is growing more complex and developing new structures, as Tok Pisin did during its expansion stage. It seems far more likely that syntactic scope and functional scope would both expand to cope with functional demand.

As with the other grammatical items analysed so far, the question in the emergence of *tasol* as an adversative conjunction is not merely where its grammaticalization path originated and which steps were part of the process, but also why alternative constructions - such as English *but* or German *aber* - were not adopted. One of the reasons might have been lexical blocking, i.e. the fact that a form similar to English *but* already existed with another meaning or function in Tok Pisin. Mihalic, for instance, records *bat* as a form in his Tok Pisin dictionary (Mihalic 1957, 12) deriving from English *bad*. In his examples - *batpasin* 'bad habit, vice' and *batpilai* 'indecent acts' - it seems to function

as a prefix. Theoretically, if it predates the necessity for an adversative conjunction, it has acted as an inhibiting factor for the adoption of *but* due to its homophony. However, *bat* on its own has only three attestations in the Z'graggen corpus. 185 tokens start with *bat-*, but 106 of these are *bataflai*, which is unrelated. Of the remaining, 23 are *bata* (< E *butter*) and nine are proper nouns. Neither *batpilai* nor *batpasin* occur. Given that there are 12,487 attestations for *nogut*, we can safely assume that *bat* 'bad' has not become productive and widespread. It is therefore unlikely to have been a major inhibitor for the adoption of *but* as a conjunction.

A similar argument can be put forward for Tolai *ba*. In Tolai, it serves as both a conjunction 'if, when' and as an adverb 'down' (Franklin et al. 1974, 118). Since there are no traces of it having been transferred into Tok Pisin - 'down' is either *daun* or *tamblo* (from *down below*, and the role of *if* as a conjunction or subjunction is served by *sapos* - it seems unlikely that the form blocked *but*, especially given the fact that conjunctions, as we have seen, emerge rather late in the development of Tok Pisin.

We might also note, however, that according to Dutton (1999, 30), some younger speakers are switching to *bat*, possibly due to the influence of renewed contact with English. Smith (2002b, 187) similarly reports that in his corpus, "the use of ***bat*** in a way apparently identical to that of ***tasol*** was frequently attested, with 59 tokens", citing, amongst others, the following example[60]:

(74) Skul laif em gutpla **bat** tu em skin i   dai tumas.
      School life 3SG good   CONJ also 3SG flesh PM tired much
      School life is good, but it's also boring.

This seems in line with Matras' observations above, in that the adversative connective is being renewed/borrowed faster than the other two connectives. Of course, the other two connectives are, on the surface, very close to their English counterparts and the difference is intransparent to speakers. It is therefore hard to say whether they would also be replaced if their structure was significantly different to *and* and *or*.

The fact that *tasol* refers to preceding syntactic material is likely tied to its origin in the expression *Em tasol*, which would be used at the end of a narrative sequence to signal that the preceding sequence was complete. This direction is further supported by substrate influence, given that Tolai *ka* 'only' (see above) also refers to preceding material.

---

[60]Translation given by Smith, glossing mine.

6.6.3 Disjunctive conjunction: *o*

The last connective this analysis will concern itself with is the disjunctive conjunction *o*. As was the case for the previous two connectives, its origin once again may seem obvious in a direct borrowing from English *or*, requiring hardly any phonological adaptation. However, I posit that once again, the entire picture is more complicated than it first presents itself, which, taken together with what is detailed for *na* and *tasol* above, allows us to draw some further conclusions about the grammaticalization of connectives in Tok Pisin.

As previously, we will first take a closer look at the modern usage of *o* before delving into its diachronic origin. *o* has three functions in modern Tok Pisin. First, it serves as the disjunctive verbal or noun phrase conjunction, as in examples 75 and 76 below:

(75) Man i   stilim kaikai **o**   brukim haus **o**   sutim pik?
     Man PM steal food   CONJ break   house CONJ shoot pig

     Did the man steal food or break the house or shoot a pig? (ZC 00304)

(76) Mi no   save em man **o**   meri   em stori.
     1SG NEG know 3SG man CONJ woman 3SG story

     I don't know whether it was a man or a woman, they tell the story.
     (ZC 02499)

As with *na* and *tasol*, the position and function of the disjunctive conjunction mirrors the position and function of English *or* entirely: *o* is inserted directly between the constituents it connects and serves to establish the idea that either the first or the second option can be true, but not both.

Secondly, *o* can serve the function of a tag in interrogative sentences, as illustrated in examples 77 and 78 below:

(77) Na   Ruamas i   askim gen Bumari, Hei yu   gat pikinini tu
     CONJ Ruamas PM ask   again Bumari Hey 2SG have child   also
     o?
     TAG

     And Ruamas asked Bumari again, Hey, you have a child too, don't you? (ZC 00325)

(78) Yes man i   askim meri, em i   askim, Yu orait   **o**?
     Yes man PM ask   woman 3SG PM ask   2SG alright TAG

     Yes, the man asked the woman, he asked, Are you all-right, or what?
     (ZC 03704)

As Verhaar (1995, 55) notes, "the tag *o?* with level intonation at the end expresses openness to an answer not matching bias, if any, with the question". In other words, it elides an additional part of the sentence, which, in the case of the above examples, would read, respectively *Hei yu gat pikinini tu o yu no gat pikinini?* 'Hey, do you have children or do you not have children?' and *Yu orait o yu no orait?* 'Are you alright or are you not alright?'. A similar function is available to speakers of English, who can end their sentences with the tag question *or not?*, expressing the same openness to an answer not matching bias that Verhaar describes for Tok Pisin.

So far, Tok Pisin *o* seems to align well with English *or*, and we have no indication as of yet that it would not be a direct borrowing. Neither will this origin be challenged by taking a look at the disjunctive conjunction in the APiCS data, as summarized below:

Language	Disjunctive conjunction	Source
Early Sranan	*efu/ofi/efi*[61]	???
Sranan	*efu/efi*[62]	???
Saramaccan	*(ée) náá (sɔ)*[63]	???
Nengee	*ofu/efu*[64]	???
Creolese	*o/or/ar*[65]	English *or*
Trinidad Endlish Creole	???	???
Vincentian Creole	*aa*[66]	???
Jamaican	-	-
Belizean	*o*[67]	English *or*
San Andreas Creole English	*or*[68]	English *or*
Nicaraguan Creole English	*or*[69]	English *or*
Bahamian Creole	*or*[70]	English *or*
Gullah	*or*[71]	English *or*

---

[61](van den Berg, Margot C. and Smith 2013, 11)
[62](van den Berg, Margot C. and Smith 2013, 11)
[63](McWhorter and Good 2012, 151)
[64](Goury and Migge 2003, 154)
[65](Devonish and Thompson 2010, 58f.)
[66](Prescod 2013, 77)
[67](Escure 2013, 98)
[68](Bartens 2013c, 111)
[69](Bartens 2013b, 124)
[70]Own knowledge
[71](Klein 2013, 145)

African American English	*or*[72]	English *or*
Krio	???	English *or*
Ghanaian Pidgin English	ɔ[73]	English *or*
Nigerian Pidgin	ɔ[74]	English *or*
Cameroon Pidgin English	*o*[75]	English *or*
Pichi	ɔ[76]	English *or* or Spanish *o*
Chinese Pidgin English	unmarked/juxtaposition[77]	-
Singlish	*or*[78]	English *or*
Tok Pisin	*o*[79]	English *no*
Bislama	*(n)o*[80]	English *no*
Norf'k	???	???
Kriol	*o*[81]	English *or*
Hawai'i Creole	ɔ[82]	English *or*

Table 6.17: Disjunctive verbal conjunctions in the 26 English-lexified languages in the APiCS data

For once, Tok Pisins column in the table above lines up well with the other languages, including Bislama. In Bislama[83], the disjunctive conjunction occurs either in the form of *o* or that of *no*:

(79)　Bae　yu　kam　**no**　　bae　yu　stap long　　haos?
　　　FUT 2SG come CONJ FUT 2SG stay PREP-L house

　　　Will you come or will you stay at home? (Crowley 2004, 175)

(80)　Bae　yu　kam　tedei　**o**　　tumora?
　　　FUT 2SG come today CONJ tomorrow

　　　Will you come today or tomorrow? (Crowley 2004, 175)

---

[72] Own knowledge
[73] (Huber 2013, 174)
[74] (Faraclas 2013, 183)
[75] (Bellama 1983, 89)
[76] (Yakpo 2009, 201)
[77] (Matthews and Li 2013, 211)
[78] Lim and Ansaldo (2013)
[79] (Geoff P. Smith and Jeff Siegel 2013, 220)
[80] (Crowley 2004, 175)
[81] (Schultze-Berndt et al. 2013, 249)
[82] (Velupillai 2013, 259)
[83] For both examples 79 and 80, translation from original source and glossing mine.

In Solomon Islands Pijin, the disjunctive conjunction matches that of Tok Pisin and Bislama as well (Regional Assistance Mission to Solomon Islands 2011, 29):

(81) Jon hem i   ranawe **o**   hem i   kam?
      Jon 3SG PM run-away CONJ 3SG PM come
      Did John run away or come?

(82) Mere i   kukim pigpig **o**   kokorako?
      Women PM cook   pig   CONJ chicken
      Are the women cooking the pig or the chicken?

Here, for once, we seem to have found a case in which Tok Pisin does match the common pattern among the English-lexified contact languages, which would be to adopt a form of English *or* as their disjunctive conjunction. In regards to its origin, it may still seem obvious at a first glance for it to be a reduced form of English *or*. Indeed, the phonological path from /ɔːʳ/ or the weak forms /əʳ/ or /ɚ/ to /o/ is immediately evident, especially given Tok Pisin's reduced vowel inventory. Such an impression is strengthened by the observation of numerous instances in which the disjunctive coordination employed in a contact language actually does derive from English *or*, as can be seen in Table 6.17 above. This would mean that while both the adversative and coordinative conjunction are examples of reanalysis and innovation within the system of Tok Pisin, the disjunctive conjunction would have been directly borrowed from English. This seems unlikely for several reasons.

For one, Ramat and Maru (2011, 657), building on the arguments put forward by Matras (1998), suggest that in "normal" circumstances (i.e., grammaticalization processes outside the scope of extensive language contact), adversative connectives have a faster cycle of renewal and a higher degree of intra-linguistic variation due to their expressive potential. Conjunctive and disjunctive connectives, on the other hand, are said to be more stable. In bilingual situations, "languages replacing combination markers also replace alternative (disjunctive) markers, and languages replacing alternative markers also replace contrast(adversative) markers" (Ramat and Mauri 2011, 657). Transferring this notion from renewal to innovation - a first step in the renewal cycle - would indicate that it is unlikely for the adversative conjunction *tasol* to have been innovated, but not the disjunctive one. Of course, one might argue that the notion might not be transferable from renewal to innovation in the first place. We should, however, consider that except for Bislama and Tok

Pisin, all of the language in tables 6.15, 6.16 and 6.17 have the same source language for all three types of connectives. This would at least seem to indicate the reverse tendency - if one connective is directly borrowed, so are the others - although the causal direction is not apparent (it might be that in case of creole genesis, the disjunctive connective is borrowed first, followed by the coordinating connective and finally, the adversative connective).

Secondly, as was the case for both *na* and *tasol* (as a conjunction), the timing is contraindicative for a direct borrowing. Like the other two conjunctions, *o* does not appear in early texts, which would have been written/recorded at a time when exposure to English as an L1 would have been greater. Neither would there have been a need for a verbal disjunctive coordination, given that sentence structures were far less complex at the time. Like the other two conjunctions, *o* only begins to appear once sentence construction has become more complex, but English as a model is no longer readily available. Given that the form *o*, or phonologically similar forms, were not preserved in the meantime in another function and therefore not present in expanding Tok Pisin, we also find no starting point from which it could have been reanalysed to a disjunctive conjunction.

We might be tempted to look for its origin in either the other superstrate, German, or in one of substrate languages next. German provides us with *oder* as a possible source form, which would require extensive phonological reduction to eventually develop into the simple form *o*. Furthermore, it would, to my knowledge, be the only grammatically functional item to be borrowed from German into Tok Pisin. All other borrowings are of purely lexical nature. It also seems unlikely that two of three connectives would be innovated, while the last one would be borrowed. In addition, as mentioned above, it would not be in line with Matras' hierarchy for the disjunctive connective to fall out of line in this regard. As thus, while the possibility for *o* to come from German *oder* cannot entirely be disregarded, it seems unlikely.

You will also have noted that for some of the languages in table 6.17 above, Spanish *o* is given as a possible source of the disjunctive connective. The same indicators that make a German borrowing unlikely largely discredit a Spanish origin: *o* only occurs pretty late into the development of Tok Pisin, when global sailor jargon would not have played a large role anymore. In addition, it is unlikely for a pattern that requires an intermediate level of complexity such as the disjunctive connective to be borrowed from simplified jargon.

This leaves us with the Austronesian and Oceanic substrates. For Tolai, Franklin et al. (1974, 111) record only 14 total lexemes with initial o: *oaga* 'canoe', *oao* 'hot, feverish', *obo* 'to revenge', *oe* 'to plant', *ogoe* 'to beat a native

drum', *oka/oko* 'to poke', *okang* 'to pick', *olo* 'to enter', *ong* 'to put something into an aperture', *ongo* 'to obey', *ongor* 'strong; to work hard', *oro* 'to call out', *oroi* 'to cease (rain)' and *ot* 'completely'. Furthermore, in the section on particles, none is recorded to start with *o*. We can thus rule out formal supply from Tolai. Neither does functional supply seem possible: while Tolai seems to have a coordinative conjunction in *ma*, the functions of the adversative conjunction and the disjunctive conjunction seem, to the best of my knowledge, not be present in any overt form.

If Tok Pisin *o* is not English *or*, likely not German *oder* and does not come from the substrate languages, then where does the form originate? I suggest that it is a reduced form not of *or*, but of *no*, the negation marker. There are several indications that this might be the case. For instance, in Mihalic's 1957 dictionary, there is no distinction in form between the negation marker *no* and the disjunctive conjunction *or*. As thus, the second entry under *no* reads as follows (Mihalic 1957, 94):

> **no** (See also: *yes.*)
>
> 1) not
>
> *Em i-no inap.* = That is not enough.
>
> *Mi no laik.* = I don't want to. I won't.
>
> *no laikim tru* = to detest
>
> 2) or
>
> *Sip i-go pinis no nogat?* = Did the ship go? Has the ship left?
>
> In English we do not usually add the "or not"; in Neo-Melanesian it is usually added.
>
> *Yu laikim dispela no arapela?* = Do you like this one or that?

Correspondingly, the entry for *o* does not yet indicate a conjunctive function. Instead, it lists three forms of usage: that as an interjection, as in English "Oh!"; that of a suffix that adds no meaning, but makes words carry over a distance; and as occurring "in Neo-Melanesian songs as well as amongst boys working in groups and calling out to one another" (Mihalic 1957, 94). Neither does Mihalic indicate that at the time of his observations, *no* and *o* would be variants of the same form. The same is true for the vocabulary section in Hall (1943a, 111), which only lists the meaning 'oh' for *o*. For *no* itself, Hall (1943a, 111) lists three entries, namely negative adverbial 'not', which includes an introductory adverbial function 'nor, and not', and minor clause 'no'. Verhaar (1995, 39), however, notes that "[...] in some older styles of Tok Pisin *no* may also mean 'or' - its more usual form is o." A conflation of *no* and

o is, however, certainly the case in modern Bislama (see examples 79 and 80 by Crowley above).

The construction of *no nogat* as a tag question has endured in Tok Pisin, although it now, naturally, occurs as *o nogat*. With 401 occurrences in the Z'graggen corpus, *nogat* is the 11th most frequent 1R collocate of *o*.

No as a negation particle in Tok Pisin predates *o* as a disjunctive conjunction in Tok Pisin. As noted above, the latter does not appear before the expansion stage, while the former is attested as early as the 1880s (Mühlhäusler 2003, 40). It is therefore necessary to point out that should Tok Pisin *o* actually derive from *no*, it would not be a borrowing of the form of English negation particle, but a functional extension or reanalysis of Tok Pisin's *no*. In other words, *no* was first borrowed as a negation particle and then expanded its function to that of the disjunctive conjunction:

no(t) $\Rightarrow$ English negation particle
$\Downarrow$
no $\Rightarrow$ Tok Pisin negation particle
$\Downarrow$
no $\Rightarrow$ TP negation particle, part of tag question
$\Downarrow$
no $\Rightarrow$ TP negation particle, part of tag question, disjunctive conjunction
$\Downarrow$
o $\Rightarrow$ TP disjunctive conjunction
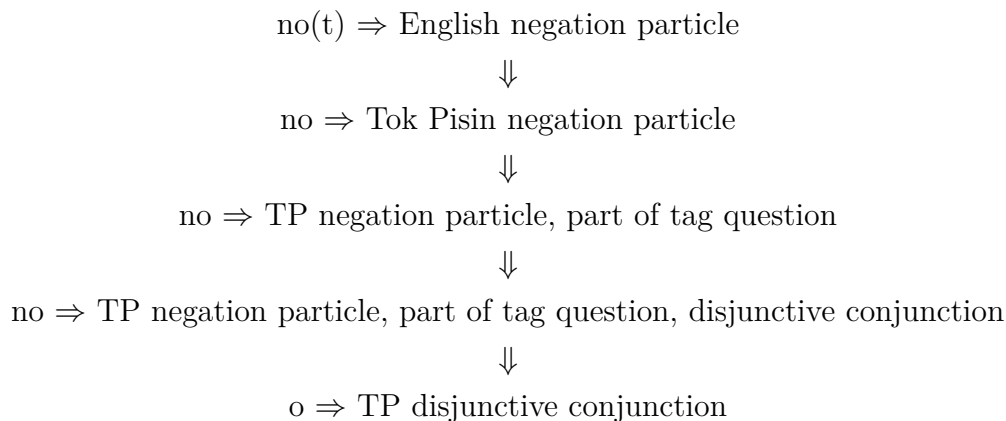
Such a development of a negation particle into a disjunctive conjunction is not unheard of. Ramat and Maru (2011, 659) suggest that it is one viable grammaticalization pathway, next to others such as distal elements, interrogative particles, free choice verbs, dubitative particles and denied conditional clauses. It should be noted, however, that they only provide a single example of negative particles becoming disjunctive connectives, namely that of Nakanai (Oceanic) *ka* deriving from the negative particle *(ou)ka*. According to their analysis, all the paths except for the distal elements "instantiate a further inherent semantic property of disjunction, namely the irrealis potential status of the two alternatives, which cannot be presented as facts, but need to be overtly indicated as possibilities" (2011, 659). Negative markers, they specify, "develop into disjunctive connectives in contexts where one of the two alternatives is overtly denied in order for the second one to be proposed".

Both the examples cited in the excerpt from Mihalic's dictionary above provide such a context or interpretation. In the first, the ship cannot have

both left and stayed, so one of the possibilities is denied so the other can be proposed. In fact, it is the underlying principle of tag questions such as *no nogat* or *isn't it/hasn't it/aren't you*/etc. to provide this function. In the second example, it would of course be possible for the interlocutor to declare that they like both options. But the very fact that the question is posed implies that the speaker wants to discard one in favour of the other.

The two examples below show that the Z'graggen corpus contains examples - although they are rare, given that during the time of its recording, *o* had already been well established - in which *no* could be parsed both as a negator and as a disjunctive conjunction. To confirm that this construction was no longer productive during the time of recording, I drew a sample of 5,000 instances of *no*. Among these 5,000 instances, there is not a single additional example of ambiguous *no* between a negator and a disjunctive conjunction.

(83) Yu lukim na   yu ting em i   no   kaikai, **no**   em
      2SG see   CONJ 2SG think 3SG PM NEG eat   NEG/CONJ 3SG
      i   kaikai.
      PM eat

      You see and you think he does not eat, (1) no, he eats. (2) or he eats.
      (ZC 01832)

(84) Ol i   laik putim tamiok i   laik brukim, **no**   stik
      3PL PM want put   axe   PM want break   NEG/CONJ stick
      bilong   tamiok i   bruk.
      PREP-P axe   PM break

      They want to put down the axe, want to break it, (1) no, the handle
      of the axe broke. (2) or the handle of the axe broke. (ZC 00028)

Bailey (2013, iii) notes that in examples like these, "these particles are best analysed as disjunctive elements, heading an elided clause". In Tok Pisin, there is, of course, no specific question particle. However, *o nogat* serves a similar function, and, if Mihalic's account is to be believed, so did *no nogat* before.

Note that it seems equally plausible for a disjunctive function of *no* to arise in a clausal scope than in one of a NP. *Sip i-go pinis no nogat* 'has the boat left **or (has the ship) not (left)**' is no more natural a context for negation, with one part of the clause being denied and the other option being expressly stated by *no nogat* 'not not have', than the nominal version. In constructions such as *yu laik dispela no arapela*, on the other hand, there is no additional syntactic material to overtly mark the second option. However, given that *no* in Tok Pisin generally negates the following phrase, it could be parsed as 'do you want this one, not the other'.

In addition, *o nogat* is still quite productive in the same position and function that Mihalic's reported *no nogat* used to be, with 140 attestations in the Z'graggen corpus, two of which will serve as examples below:

(85) Na   em askim, Sori brata mi sut   long   pisin na
      CONJ 3SG ask   Sorry brother 1SG shoot PREP bird CONJ

      banara i   kam stret hia, yu   lukim **o nogat**?.
      and   arrow PM come street here 2SG   see   TAG NEG

      And he asked, Sorry brother, I shot a bird and the arrow came to this
      street, did you see it or not? (ZC 00910)

(86) Em i   stori bilong   yu tu **o nogat**?
      3SG PM story PREP-P 2SG also TAG   NEG

      Is this also your story or not? (ZC 00139)

From this, we can conclude that the earlier tag question *no nogat* eventually became *o nogat*, likely either for reasons of ease of articulation or to prevent confusion between the initial *no* and the first syllable of *nogat*.

6.6.4 Summary

To summarize, I have suggested that none of the three connectives *na*, *tasol* and *o* was directly adopted from English in their function. Rather, they have been grammaticalized and/or reanalyzed from forms that were previously brought into Tok Pisin to serve a different function. In the case of *na*, it was the adverb *nao*, while *tasol* first occurred as a discourse marker and *o* has its origin in the negation marker *no*. None of these cases involved processes previously unknown or leaps of creativity, and neither did they involve much agency by the speakers. Rather, step by step, they can be accounted for by well-documented processes of linguistic change.

On the other hand, the developments sketched above serve as a cautionary tale against assuming the easiest or most obvious path in language contact. Just because the form looks similar in the contact language and one of its superstrates, that does not mean that the process by which it came about was one of simple borrowing. Furthermore, for the connectives, the same holds true as for several of the developments described previously: the time period during which they came about and the resulting extralinguistic factors likely heavily influenced their development. Had English L1 input been more readily available, it would have been quite likely that the English forms would have

been imported along with their function.

6.7 Concessive: *maski*

In modern Tok Pisin, *maski* has at least three distinct functions. It can be either a concessive conjunction, a prohibitive marker or a sentence adverb. In the following, each of the functions is described briefly. A concessive conjunction introduces a phrase which could conceivably preclude the event described in the main clause, but does not. In English, this function is served by constructions such as *even though, although* or *despite.* Verhaar (1995, 439) notes that In Tok Pisin, "as a conjunction, [*maski*] is concessive, and may also take the form *maski sapos* if the concessive part is conditional". He sees it as "not quite parallel [to] English *although*", commenting that there are two notable differences. First, in a Tok Pisin subclause *maski* indicates that the information carried by the sentence is somewhat irrelevant to the information carried in the higher clause. This is, naturally, not the case in English. In sentences such as *Although it was raining, we still went for a walk*, the information carried by the lower clause - namely, the presence of rain - is certainly relevant. Second, *although* is not necessarily conditional, but concessive *maski* always is. In other words, *maski* is always future-oriented - it would occur in the Tok Pisin equivalents of sentences such as *Even if you turn out to be right, I still won't care*, but not in sentences such as *Even though you did turn out to be right, I still think it was the right thing to do. Maski* thus carries a conditional element; this will be relevant in establishing the grammaticalization path of *maski* later on. The fact remains, however, that *maski* occurs as a concessive with a conditional element, never as a pure conditional - which is why it can be combined with *sapos* in the first place, *maski* expressing the concessive element and *sapos* expressing the conditional element of the construction. While there are no occurrences in the Z'graggen corpus for *maski sapos* - which is hardly surprising, given that there are only 77 occurences of *sapos* altogether - there are plenty of instances of *maski* serving as a concessive:

(87) Em tok, **Maski**   yu lusim, mi   mas kaikai.
      He say, CONJ-CONC 2SG leave, 1SG must eat.
      He said, even if you leave, I must eat. (ZC 02533)

(88) Orait em tokim ol, **Maski** yupela slip antap long
Alright 3SG tell 3PL, CONJ-CONC 2PL sleep on PREP
bet, bai mi slip aninit long bet long sit bilong
bed, FUT 1SG sleep underneath PREP bed PREP-L seat PREP-P
paia.
fire.

Alright he told them, even if you sleep on top of the bed, I will sleep underneath the bed in the ashes. (ZC 02778)

(89) **Maski** em i go sutim wanpela bikpela pis long
CONJ-CONC 3SG PM go shoot one big fish PREP-L
Ungu, em wara bilong mipela.
Ungu, 3SG water PREP-P 1PL-EXCL

Even if he goes shoot a big fish at Ungi, it is our water. (ZC 03541)

(90) Na **maski** supos mama i paitim yu
CONJ CONJ-CONC CONJ-CONC mother PM hit 2-SG
rausim yu, yu mas krai, yu strong na yu bihainim.
throw-out 2-SG, 2SG must cry, 2SG strong CONJ 2SG obey.

And even if your mother hits you throws you out, you must cry, you are strong and you obey. (ZC 03801)

In addition to being used as a concessive conjunction, *maski* can also serve as a prohibitive marker, i.e. a marker establishing a negative imperative for the following verb, as in "*don't* talk". This is fairly unusual in that among most of the English-lexified contact languages surveyed in the APICs data, the sole negator in both the positive and negative imperative constructions is a derivative of *no*, as table 6.18 shows:

Language	Type of prohibitive	Negator
Early Sranan	normal imperative construction, normal negator	*no*
Sranan	normal imperative construction, normal negator	*no*
Saramaccan	normal imperative construction, normal negator	*naa*
Nengee	normal imperative construction, normal negator	*na*
Creolese	normal imperative construction, normal negator	*na*
Trinidad Endlish Creole	normal imperative construction, normal negator	*doh*
Vincentian Creole	normal imperative construction, normal negator	*na*
Jamaican	normal imperative construction, normal negator	*no*
Belizean	special imperative construction, normal negator (70%); normal imperative construction, normal negator (30%)	*no*
San Andreas Creole English	normal imperative construction, normal negator	*no*

Nicaraguan Creole English	normal imperative construction, normal negator	*no*
Bahamian Creole	normal imperative construction, normal negator	*don't*
Gullah	normal imperative construction, normal negator	*don*
African American English	normal imperative construction, normal negator	*don't*
Krio	normal imperative construction, normal negator	*no*
Ghanaian Pidgin English	normal imperative construction, normal negator	*no*
Nigerian Pidgin	normal imperative construction, normal negator	*no*
Cameroon Pidgin English	normal imperative construction, normal negator	*no*
Pichi	special imperative construction, normal negator (50%); normal imperative construction, normal negator (50%)	*no*
Chinese Pidgin English	normal imperative construction, normal negator	*no*
Singlish	normal imperative construction, normal negator	*don't*
Tok Pisin	normal imperative construction, special negator (90%), normal imperative construction, normal negator (10%)	*maski, no, no ken*
Bislama	normal imperative construction, normal negator	*no*
Norf'k	normal imperative construction, special negator	*dana, dunt, noe*
Kriol	normal imperative construction, normal negator (70%); normal imperative construction and special negator (30%)	*nomo, don*
Hawai'i Creole	normal imperative construction, normal negator	*no*

Table 6.18: Prohibitive systems in the 26 English-lexified languages in the APiCS data

While Smith and Siegel (2002b) note that *no* does occur in the prohibitive function, "much more common is the use of the negated permissive marker *no ken* [...] or the use of the prohibitive *maski*. As example 93 below shows, some speakers will use both *noken* and *maski* in the same sentence:

(91) Na   bikpela   bilong   em tok, Ah sarap, **maski** long
      CONJ big-brother PREP-P 3SG say Ah shut up   PROH
      toktok, helpim em, yumi   yet   yumi   i   mas helpim
      PROH talk   help 3PL-INCL REFL 3PL-INCL must help body.

bodi.

And his big brother said, ah shut up, do not talk, help her, we ourselves
we must help the body. (ZC 01383)

(92) **Maski** long   em i   hambak.
PROH PROH 3SG PM misbehave

He must not misbehave. (ZC 02548)

(93) Na   lapun mama tok, Maski, **maski**, **noken** sutim em.
CONJ old   mother say, PROH PROH PROH shoot 3SG.

And the older mother said, don't, don't, don't shoot him. (ZC 00030)

(94) Ol   tok, Ah **maski** rausim   pikinini bilong   yu [...].
3PL say Ah PROH throw-out child   PREP-P 2SG

They say, Ah don't throw out your children [...]. (ZC 02306)

> Negative imperatives, or prohibitives as we shall call them, are formed with
> *nogut* or *maski*. Nogut precedes the verb and requires a subject in between
> *nogut* and the verb. By contrast, after *maski* (which may be followed by long)
> as a marker of the prohibitive, a subject is not needed. (Verhaar 1995, 43)

In Verhaar (1995, 407) we find the remark that "as a sentence adverb, *maski*
is followed by a pause". As such, he points out, *maski* [pause] *rausim* would
translate into English as 'never mind, throw it out', whereas *maski rausim*
without a pause would indicate a prohibitive 'don't throw it out'.

(95) Na   tupela tingting, Ah **maski**, mitupela   traim tanim ol
CONJ DT   think,   Ah ADV   1PL-INCL try   turn PLM
pisin.
bird

And both thought, Ah doesn't matter, we try to turn into birds. (ZC
01340)

(96) Na   mama i   tokim pikinini bilong   em, Ah **maski**, ol
CONJ mother PM tell   children PREP-P 3SG Ah ADV   PRN
i   ken i   kam kros.
PM can PM come angry

And the mother told her children, ah doesn't matter, they can get
angry. (ZC 01930)

(97) Na   em tok, Oh **maski**, yutupela was   long   pik, mi bai i
CONJ 3G say Oh ADV   2PL   watch PREP pig 1SG FUT PM
go.
go

And he said, oh doesn't matter, you watch the pig, I will go. (ZC
02982)

(98) Na   em tokim ol, Ah i   orait **maski**, nau   yupela stap,
      CONJ 3SG tell   3PL Ah PM alright ADV   ADV-T 2PL   stay,

      bai mi go na   traim was.
      FUT 1SG go CONJ try   watch

      And he told them, ah it's alright doesn't matter, now you stay, I will
      go and try to watch. (ZC 03817)

The origin of the form *maski* is somewhat disputed in the literature. Mühlhäusler
(2003, 45) remarks that "*maski* 'nevermind' is a word widely found in the pidgins
and creoles of the world". Rusling (1874, 303) thus reports the following:

> *Chop-chop* means 'very fast'; *maskee*, 'don't mind', *Topside Galah*, 'Excelsior,
> hurrah!'. If you call on a lady and inquire of her Chinese servant "Misee have
> got?" He will reply, if she be up and about, "Missee hab got topside"; or, if she
> be still asleep "Missee hab got; wakee sleepee." Not wishing to disturb her, you
> hand him your card, and go away with, "Maskee, maskee; no makee bobbery!"

It is often alleged to be of Portugese origin, but sometimes German *macht
nichts* 'nevermind' is said to have reinforced its use in Tok Pisin. Engelberg and
Stolberg (2017, 38f.) make a phonological argument against a direct borrowing
of *maski* from German *macht nichts*, however:

> [None of the previous works] mention the difficulty in explaining the phonology
> of *maski* if it were derived from German *macht nichts*. While German *ch*
> [ç] can be rendered as [s] in Tok Pisin (cf. Tok Pisin *tais* < German *Teich*
> 'pond'), it seems puzzling why [t] should be turned into [k], as [t] is perfectly
> common in Tok Pisin. The apparent reduction of *nichts* to *(k)i* would also be
> at least unusual, compared to the phonological adaptation of other German-
> based items in Tok Pisin.

While it has to be pointed out that *macht* has the German *ach*-Laut [x] and
not the German *ich*-Laut [ç], the rest of the argument stands. In general, they
conclude (2017, 39f.) that "it can be said that maski, while being assigned a
German etymology in several sources, is highly unlikely to be of German origin". Instead, they as well point to the fact that it is a "well-stablished item in
various pidgin languages [...] as well as in Malay (Veiga and Fernández 2012:
197), that is, varieties with which Melanesian Pidgin English and, later, Tok
Pisin were in contact".

Mühlhäusler (1985, 210) similarly argues against overrating the influence
of Spanish and Portugese in the development in Tok Pisin. While he acknowledges
that both nations made contact with the area long before they were
English and German colonies, neither contact situations were intimate enough

or lasted long enough in order to leave substantial linguistic traces on the local languages. Certainly, at the time of Tok Pisin's stabilization and expansion, contact with neither Portugese nor with Spanish played a significant role. While there are traces of lexical borrowings, such as *pikinini* (Portugese *pequeno*), he notes that "all evidence examined by me suggests that these words were already established in Australian or Pacific English at the time Tok Pisin came into being, and that they were probably borrowed from a variety of English". He further acknowledges that "*maski* never mind may also have found its way into Tok Pisin via Chinese Pidgin English" (1985, 200), while cautioning that the extent of CPE's influence on Pacific English Jargon in the Melanesian area is unclear, as is the question of "how much linguistic continuity there is between these early jargons and late Tok Pisin".

In Hall's (1943a, 108) vocabulary section, *maski* is listed as having a "basic meaning 'it is a matter of indifference'", with three distinct functions: an adverbial 'as far as I am concerned; never mind'; a preposition 'in spite of'; and a minor clause 'O.K. by me'. Mihalic's (1957, 82) dictionary records two distinct meanings for *maski*, the first of which is "to be indifferent; it does not matter; no matter; who cares?". The second recorded meaning is "in spite of, despite", which suggests a prepositional or subjunctional use of *maski*. However, though none of Mihalic's examples for this second meaning are actually translated with either of these two options. Instead, he uses *never mind*, e.g. "*Maski long kukim kaikai.* Never mind cooking. Don't bother cooking anything". Whether this is an editorial error or Mihalic was unable to find more suitable examples for this second meaning, I cannot say.

However, which language it was ultimately derived from and how exactly it entered Melanesian Pidgin is not the central issued of interest here. We can reasonably assume that if it was borrowed as early as previous discussion suggest, it was borrowed neither as a concessive conjunction, nor as a prohibitive marker, but instead as a lexical item or idioms with the meaning 'nevermind', which eventually turned into a sentence adverb used as described above. From there, it would have grammaticalized to the other functions. I suggest that this development occurred as follows:

$$maski \Rightarrow \text{Lexical item/idiom 'nevermind'}$$
$$\Downarrow$$
$$maski \Rightarrow \text{Sentence adverb}$$
$$\Downarrow$$
$$maski \Rightarrow \text{Sentence adverb, conditional concessive}$$
$$\Downarrow$$
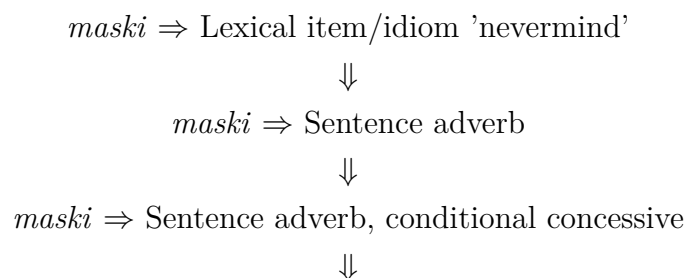
maski ⇒ Sentence adverb, conditional concessive, prohibitive

The actual process by which *maski* likely developed from a sentence adverb into a conditional concessive is one of syntactic reanalysis due to its position within the sentence. Consider the fact that *maski*, as a sentence adverb, frequently occurs within a sentence and not in final position: only 279 of the 2,877 occurrences are sentence-finally. Instead, *maski* is often positioned right before a predicate or clause even as a sentential adverb:

(99) Orait **maski**, yupela i go bek.
[...] Alright no-matter 2PL go back

Alright, it doesn't matter, you go back. (ZC 04110)

(100) Ol lukim bai ol save tok, Oh **maski**, larim em.
3PL see FUT 3PL HAB say Oh no-matter let him

They see, they will say Oh it doesn't matter, let him. (ZC 03719)

(101) [...] Orait **maski**, mi bihainim ol, yupela i stap.
[...] Alright no-matter 1SG follow 3PL 2SG PM stay

Alright, it doesn't matter, I will follow them, you stay. (ZC 00486)

(102) Na bikbrata kirap tokim em, Ah **maski**, mi bikman
CONJ big-brother INCH tell 3SG Ah no-matter 1SG big
pinis, yu wantaim meri bilong yu, yutupela ken i go
COMP 2SG PREP-C woman PREP-P 2SG 2PL can PM go
sindaun long dispela nupela haus.
lie-down PREP-L DT new house

And the big brother started telling him, Ah, it doesn't matter, I'm grown, you and your woman, you can go lie down in this new house. (ZC 03663)

Note that in examples 101 and 102, both a sentential-adverbial reading and a concessive reading are possible grammatically. The last example could just as well be parsed as 'even if I'm grown, you can [...]', as can the one before: 'even if I follow them, you stay'. What we are dealing with here is a scope expansion from a sentential adverb as a syntactically independent unit to a concessive conjunction with its scope on the following phrase. This is made possible by the fact that it does not change the final meaning of the sentence: both 'don't worry, X will do Y, Z will do N' and 'even if X does Y, Z will do N' result in the latter - and in the concessive reading, higher - clause carrying the same main information. Sentences such as this are therefore likely to have been the model for syntactic reanalysis of *maski* as a concessive conjunction. Moreover, sentences of this type are not rare.

It does not yet explain how *maski* came to be used as a prohibitive, though. In general, this is the rarest function of *maski* in the Z'graggen corpus, which is another indication that it might be the most recent. In a sample of 200 occurrences, only 3 were of the prohibitive function, while 43 functioned as concessive conjunctions and the remaining 154 were sentential adverbs.

Here, the competition between *noken* and *maski* sheds light on the origin of the latter. As pointed out above, their original function is distinctive in that *noken* is the negated form of the permissive, whereas *maski* serves as the prohibitive, i.e. *noken* indicates actions that the addressee is not allowed to perform, whereas *maski* precedes actions that the addressee should not perform [84] The fact that these are similar, but distinct functions is once again evident in both forms co-occuring in the same predicate, as below:

(103) Na   liklik brata i   tok, Oh **maski** noken   toktok, ai
       CONJ small brother PM say Oh PROH NEG-PERM talk   eye
       tasol i   lukluk.
       only PM see
       And the small brother said, Oh don't, you are not allowed to talk, only look with your eyes. (ZC 003846)

Syntactically, the two forms also behave differently: while *noken* is often part of the predicate (occurring in the constructions of *i noken i* + VB, *i noken* + VB and *noken i* + VB 471, 2,235 and 658 times, respectively), there are only three instances of *maski* being preceded by *i* and 88 occurrences of it preceding *i*. This discrepancy reflects the different origins of *maski* - a sentence adverb, i.e. its own syntactic unit - and *noken* as an auxiliary, i.e. as dependent on another predicate.

Unfortunately, it is hard to compare the usage of *noken* and *maski* quantitatively, given the high occurrence of the former on the one hand and the rare occurrence of the latter in the function of the prohibitive on the other hand. While there are some tendencies for certain verbs to occur only with one of the forms - *pilai* 'to play', for instance, only occurs with *noken*, which might be caused by playing requiring permission more often than *kilim*, which occurs more freely with both forms - but their numbers are too low to make a final determination.

The question as to how *maski* came to express this third function of the prohibitive remains, however. I would argue that reanalysis also happened

---

[84]For a detailed examination of the development of *ken* to a permissive marker, see the section on the abilitative complex in the chapter on tense, mood and aspect markers below.

in this regard, though on a different scope than with the concessive above. Consider sentences such as the one below:

(104) [...] **maski** long bikpela brata.
[...] not-worry PREP big brother.
[...] don't worry about the big brother. (ZC 00968)

Tok Pisin does offer its speakers the possibility to integrate *maski* more closely into the sentence structure than just a sentence adverb. Here, it is accompanied by *long* as a preposition, which does, of course, require a prepositional complement. The meaning becomes more specific: not 'don't worry' in general, but 'don't worry about this specific thing'. However, the complement position can not only be filled by noun phrases:

(105) Wanpela diwai tasol yu katim, **maski** long katim planti.
NUM tree only 2SG cut not-worry PREP cut many
Cut only one tree, don't worry about cutting many. (ZC 02539)

(106) **Maski**, yumi wokim bet na karim i go long
Not-worry 1PL-INCL craft shelf CONJ carry PM go PREP-D
ples, maski long painim abus.
place not-worry PREP find meat
Don't worry, we'll build the shelf and carry it to the place, don't worry about finding meat. (ZC 02949)

This complement is the predicate of the second clause, which, in English, is rendered by the gerund. The essence of the second clause of such a sentence - the type of 'don't worry about doing X' - is 'do not do X'. The sentence is perfectly parsable as 'Cut only one tree, don't cut many trees'. In this case, then, there would also be an expansion of scope, or rather, an establishment of scope from the sentence adverb to the prohibitive. While this seems a likely path, the scarcity of *maski* in this function in the available data makes it impossible, at the moment, to back it up with evidence. As such, the grammaticalization paths for *maski* appear as follows:

The processes that led *maski* to develop along these paths were, as detailed above, mostly syntactic reanalysis. With both paths and processes being established, there is one more question to ask about *maski*, namely why its functions did not fall to other forms. Heine and Kuteva (2002, 335) list only one possible source for the prohibitive in their overview, namely STOP, and two for the concessive, namely CONDITIONAL and TEMPORAL (2002, 335.

```
prohibitive — sent. adverb — concessive
                       |
                 'don't worry'
```
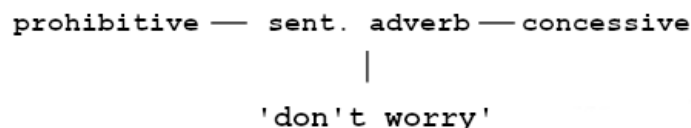
Figure 6.7: Grammaticalization paths for *maski*

Stop, however, took a different path in Tok Pisin, eventually emerging as the continuous marker as described in section 6.8.2 below.

The prime alternative for *maski* as a concessive through the TEMPORAL path would likely have been *taim* 'when'. Clauses such as *taim yu kam bek* 'when you come back' could easily have been reanalysed first as 'if you come back', then further to 'even if you come back'. However, this did not happen: *taim* has no conditional aspect to it in Tok Pisin. Neither did *sapos*, a conditional marker, extend to the concessive, likely due to its low frequency. However, the concessive still arose through the conditional path, more or less: *maski*, as a concessive conjunction, has an obligatory conditional element and is future-oriented (relative to the moment of speech). This is also evident in the fact that in the 2,877 occurrences of *maski* in the Z'graggen corpus, *bai* occurs within an environment of 5L and 5R 425 times, while *bin* occurs only 3 times.

6.8 Tense-Mood-Aspect markers

Tense, mood and aspect are encoded by pre- or postverbal markers in modern Tok Pisin. Most of these derive from an English etymon, with the exception of the habitual marker *save*, which derives from Portugese *saber*. The following table gives an overview of the TMA markers Geoff P. Smith and Jeff Siegel (2013) list for Tok Pisin:

form and position	function	etymon
(i) bin V	past tense	*been*
bai (i) V	future/irrealis	*by and by*
V *pinis*	completive	*finish*
(i) sa(ve) V	habitual	Portugese *saber*
(i) wok lo(ng) V	continuous	*work along*
V *i stap*	continuous	*stop*

(i) mas V	obligative	*must*
(i) ken V	permissive	*can*
(i) (i)nap V	abilitative	*enough*

Table 6.19: Tense-mood-aspect markers in Tok Pisin

These TMA markers are used in the following order in the verb phrase (quoted from Geoff P. Smith and Jeff Siegel (2013) as adapted from Sankoff (2010)):

bin				i go		
no	save	laik	V	i stap	pinis	
ken				i kam		
	ken					i kam

Table 6.20: TMA markers in the Tok Pisin verb phrase

Of these markers, several are of interest to the present study due to their origin, their function and/or their grammaticalization process. These are completive *pinis*, abilitative *(i)nap* and the two variants of the progressive marker, *wok lo(ng)* and *i stap*. In addition, *kirap* and *kamap* can be seen as inchoative markers, a development that will be adressed below.

6.8.1 The past complex: *bin* and *pinis*

Bin serves as an anterior marker in Tok Pisin, i.e. it establishes a relative past tense. In order to adequately analyze it, it will first be necessary to distinguish between the closely related concepts of *past* markers and *anteriority* markers. The most common distinction between the two is that the former marks an absolute tense, while the latter marks a relative tense. In relation to anteriority, Hackert (2004, 15) notes the following:

> Anteriority, for example, [...] refers to a particular semantic feature characterizing TMA categories such as the PAST or the PERFECT rather than to a category itself. [T]he label "anterior" [...] emphasizes the relative nature of creole past markers, whose meaning may have to be rendered either by the [Standard English] Simple Past or the Past Perfect [...]; "[t]his early characterization, however, soon gave way to the simplistic and inaccurate notion that 'anterior' expresses 'simple past' with statives and 'past before past' with nonstatives."

The problem with this "simplistic and inaccurate notion" - the relativity or absoluteness of creole past markers relying on whether they precede stative or nonstative verbs - is encountered in creoles which use an anterior/past marker for relative contexts, but leave the verb unmarked in absolute contexts, regardless of whether the verb is stative or nonstative. This is the case in Tok Pisin. Here, *bin* marks both "a past-before-the past" (Verhaar 1995, 313) and "a point in time anterior to any time that is, anterior to nonpast" (present or future). Both of these usages are relative, i.e. anterior. The fact that *bin* can mark points in time which are anterior to a point in the future, but still posterior to the present disqualifies it as a past marker.

The notion that stative verbs combined with anterior markers express simple past, while nonstatives are 'past before past' is easily disproven for Tok Pisin in two ways. First, stative verbs occur with the anterior marker in contexts where it is obvious that a relative tense relation is in place:

(107) Na   han bilong diwai kokomo i   **bin** sindaun long
       CONJ hand PREP-P tree   hornbill PM ANT sit-down PREP-L
       em na   noisim   ia i stap yet, em i no   bruk.
       3SG CONJ make-noise FOC PM PROG still 3SG PM NEG break
       (ZC 02790)

       And the branch the hornbill had sat down on and was making noise,
       it did not break.

Secondly, nonstative verbs occur not only with 'past before past', but also with what Verhaar calls 'anterior to nonpast'. Nonpast, in this definition, includes both present and future as well as hypothetical past (or counterfactual) (1995, 314). In other words, *bin* is a relative past marker - or anterior marker. As Verhaar points out, "[w]hen *bin* marks anterior to past, it often corresponds to the English pluperfect tense", and "when bin marks anterior to the present time, then it corresponds with the English perfect tense in that "the anterior event or state[85]" still influences the present (1995, 315).

As such, Tok Pisin differs greatly here from English, which has grammaticalized various means of referring to the past, including for the pluperfect and the perfect tense. In Tok Pisin, both of these are rendered by the use of *bin*. In contrast, any unmarked sentence in Tok Pisin can be parsed as either present or simple past. For instance, *Mi go long haus* could both mean 'I went to the house' or 'I'm going to the house'. To circumvent such ambiguity,

---

[85]Notice how Verhaar specifically includes stative verbs.

speakers of Tok Pisin can indicate absolute past through temporal adverbials or prepositions (see the third and fourth example below).

(108) Bipo   ol i **bin** kam sindaun long Mafor.
       PREP-T 3PL PM ANT come PREP-D Mafor

       Before they had come to Mafor. (ZC 04176)

(109) Husait i **bin** stori,   husait i bin stori long yu.
       Who   PM ANT tell-story who   PM ANT tell stori PREP-D 2SG

       Who told the story, who told the story to you. (ZC 03854)

(110) **Aste**   yu krai long   kapiak.
       Yesterday 2SG cry PREP breadfruit

       Yesterday, you cried for breadfruit. (ZC 03086)

(111) Na **bipo bipo** tru mipela   i   no   save long   God.
       CONJ before before true 1PL-EXCL PM NEG know PREP God

       And before, long before, we did not know God. (ZC 03918)

In this survey of tense-mood-aspect markers and their origin, *bin* is one of the least interesting. That is not to say that it does not have intriguing features and unanswered questions as to its usage. It is, however, formally rather ordinary, as the following table shows:

Language	Past/anteriority marker	Future/irrealis marker
Early Sranan	*ben*	*de go*
Sranan	*ben*	*o*
Saramaccan	*bi*	*o*
Nengee	*be*	*o*
Creolese	*bin*	*sa, gu, a gu, gain*
Trinidad Endlish Creole	*bin*	*wud*
Vincentian Creole	*bin*	*a*
Jamaican	*wehn*	*wi*
Belizean	*me*	*wan*
San Andreas Creole English	*wehn*	*gwain*
Nicaraguan Creole English	*did/mi*	*gwain/will*
Bahamian Creole	*did*	-
Gullah	*been*	*gwine, ga*
African American English	*had/bin*	*gon*

Krio	*bin*	*go*
Ghanaian Pidgin English	-	*go*
Nigerian Pidgin	*bin*	*go*
Cameroon Pidgin English	*bin*	*go*
Pichi	*bin*	*go*
Chinese Pidgin English	*hap*	-
Singlish	*-ed*/irregular forms	???
Tok Pisin	*bin*	*bai*
Bislama	*bin*	*bae*
Norf'k	*bin*	*gwenna*
Kriol	*bin*	*-l*
Hawai'i Creole	*wen*	*goin/gonna*

Table 6.21: Past and future marking systems in the 26 English-lexified languages in the APiCS data

As we can see, the form *bin* as a past or anterior marker is not only common among English-lexified creoles, but seems to rather be the norm. Neither is its origin a mystery: it is obviously derived from English *been*. Such a grammaticalization has been attested for many creole languages (Bruyn 2009, 397). As such, it also occurs in the other two Melanesian Pidgin dialects. For Bislama, Crowley (2004, 93) points out that while many speakers use *bin* as an "ordinary past marker", some speakers use it only to "refer to [...] prior past, referring to things that happened in the past that took place before some other following event", i.e. an anterior marker as it is in Tok Pisin. The examples below show one instance of past/anterior marking with and one without *bin*[86]:

(112) Mi **bin** go.
       1SG PST go

       I went.

(113) **Yestedei** mi go long taon.
       Yesterday 1SG go PREP-D town

       I went to town yesterday.

---

[86]Glossing mine, translation as given by Crowley.

The past marker in Solomon Islands Pijin is *bin* as well, as the examples from Huebner (1979, 57) below show[87]:

(114) Alfred hem i **bin** wokabaot long bus.
Alfred 3SG PM PST walk PREP-L bush

Alfred walked/has been walking in the bush.

(115) Fred hem i **bin** wokabaot.
Fred 3SG PM PST walk

Fred walked/has been walking.

Note that *bin* only serves as an anterior *tense* marker, i.e. capable of expressing perfect and pluperfect relations (i.e. those describing past events with present relevance and past events with relevance in a more recent past, respectively), not an aspect marker encoding perfective relations (i.e. describing a completed action). While Tok Pisin has no perfective marker in the narrow sense, a related role falls to *pinis*, which Mühlhäusler (1985, 380) notes expresses "the completion of an action [...] typically [...] impl[ying] that an action is irreversible and that its result is felt for some time".

(116) Na ol i kaikai **pinis** na em i tokim ol, Yupela
CONJ 3PL PM eat COMPL CONJ 3SG PM tell 3PL 2PL
i singsing nau.
PM sing now

And they finish eating and he tells them, you sing now. (ZC 02530)

(117) Em i haitim **pinis** na giaman tok mi no lukim dispela
3SG PM hide COMPL CONJ lie tell 1SG NEG see DT
spia.
spear

He finishes hiding it and lies, tells me he has not seen this spear. (ZC 00382)

(118) Na man ia harim olsem em katim het bilong em na
CONJ man FOC hear PRN 3SG cut head PREP-P 3SG CONJ
i dai **olgeta**.
PM die COMPL

And the man hears that, he cuts off his head and he dies. (ZC 02971)

(119) Ol i trautim **olgeta pinis**, bel bilong ol i
3PL PM vomit COMPL COMPL belly PREP-P 3PL PM
slek olgeta.
become-loose entirely

They finish vomiting, their bellies become loose. (ZC 01463)

As is *pinis* in Tok Pisin, *finis* in Bislama is used in final position. See the examples below, cited from Crowley (2005), annotation mine:

(120) Tomson i   hanggre **finis**.
       Tomson AGR hungry   COMPL
       Thompson is already hungry.

(121) Ol studen oli   ridim buk **finis**.
       PL student AGR read   book COMPL
       The students have already finished reading the book.

Solomon Islands Pijin works similarly in this regard, as examples 122 (Jourdan 2008, 481) and 123 (Huebner and Horoi 1979, 62) show.

(122) Wakum gaden blong   mifela **finis**.
       Work   garden PREP-P 1PL   COMPL
       I have completed my work in our garden.

(123) Iu   kam taem san hem go daon **finis**.
       2SG come when sun 3SG go down COMPL
       Come after the sun sets.

It is worth noting that *pinis* (in Tok Pisin) and *finis* (in Bislama and Solomon Islands Pijin) can also be used as full verbs, as indicated by the examples below, fhe first of which is from Tok Pisin, followed by Bislama (Crowley 2004, 35) and Solomon Islands Pijin (Huebner and Horoi 1979, 62):

(124) Orait em i   wokim basket bilong   em, orait em i   **pinis**
       Alright 3SG PM work   basket PREP-P 3SG alright 3SG PM finish
       krai.
       cry
       Alright, she works on her basket, alright she finishes crying. (ZC 04612)

(125) Japta ia i   finis **finis**.
       Chapter FOC PM finish COMPL
       The chapter is already finished.

(126) Wea   nao bae yu go waka bihaen yu **finis** lanem Pijin?
       Where now FUT 2SG go work after   2SG finish learn Pijin
       Where will you go work after you finish learning Pijin?

Mühlhäusler (1985, 380) also notes that "[e]xamples of the use of *pinis* with non-verbal predicates are often translated by using certain adverbs or adjectives indicating completion" and that "often English uses two different lexical items to translate a Tok Pisin adjective or verb with or without the completion marker." As such, it would be more accurate to call *pinis* a completive marker rather than a perfective marker. Verhaar (1995, 316f.) points out that "*pinis* imparts even to stative predicates a prior process element" and that it "carries with it the realis modality", prohibiting its negation and making its combination with *bai* as a future/irrealis marker exceedingly rare. While the latter is true - *bai* and *pinis* occur together[88] only 659 times out of the 72,694 total occurrences of *pinis*, *pinis* and *bin* occur together even more rarely, at only 264 occurrences.

For once, neither the origin of *bin* nor the origin of *pinis* are a mystery. For the latter, Heine and Kuteva (2002, 331) list FINISH as one of the source concepts for the completive. As thus, Tok Pisin *pinis* falls right into this category. Other possibilities according to them are LEAVE, PUT and TAKE. The former was adopted directly from Pacific Jargon English and requires no grammaticalization path within the system of Tok Pisin. Neither is *bin* likely to have come about through substrate influence, given its preponderance among creoles. As for *pinis*, although Tolai has a similar construction in intransitive *par* and transitive *vapar* 'to do sth. completely', these are serial verbs tied in their positioning to the previous verb, whereas *pinis* is more free in its syntactic position as an aspectual marker. This fact, along with some differences in meaning, leads Mosel (1980, 125) to conclude that "the similarities between Tolai *par*, *vapar* and Tok Pisin *pinis* are merely accidental and cannot be explained as a result of Tolai substratum influence on Tok Pisin".

What is interesting about them is previous competition for past marking and the fact that *pinis*, like *i stap*, *i kam* and *i go*, often occurs post-verbally. Unlike them, however, it may also occur clause-finally. This seems to be a strong tendency. Romaine (1992b, 664) remarks that "despite the existence of a few examples where *pinis* occurs preverbally [...] there seems to be no trend to incorporate this aspect marker into preverbal position".
Aitchison (1989, 157) summarizes the development of *bin* and *pinis* as follows:

- Existence of two pastness markers, *bin* and *pinis*.

- Specialization of the meaning of each, *bin* to a general 'pastness' particle, *pinis* to the expression of perfectivity/anteriority, translatable as

---

[88]Defined as *pinis* occurring within the 4R collocates of *bai*.

'after'.

- Increasingly frequent use of *bin*, with considerable redundancy.

- Firm localization of *bin* between NP and verb.

- Combination of *bin* with other particles.

It is certainly the case that in the past, there was some overlap between the functions of *bin* and *pinis*. Mihalic reports both *pinis* and *bin* could be used to express completed action, with the former being "common throughout the territory" and the latter "particularly [occurring] in the Rabaul and Morobe areas". However, there are still several issues with this developmental summary. First of all, as laid down above, *bin* is by no means a general 'pastness' participle, but an anterior marker (i.e., it marks relative past), whereas *pinis* is a completive marker (i.e., it conveys an action that has been carried out completely and to its end). I also see no evidence for increasingly frequent use of *bin*, especially given its redundancy. The Z'graggen corpus holds a total of 6,210 occurrences for *bin*, outnumbered heavily already by the occurrences of *i V pinis* at 19,408. Neither is there evidence that *bin* was ever located anywhere else but between NP and the verb.

Rather than *bin* and *pinis* both emerging as pastness markers, however, I think it more likely that *pinis* emerged as a perfective/completive marker first, before coming to be used as a general past marker later on. *Bin*, in continuance of earlier Jargon English, emerged as a general past marker first, before expanding to the anterior. For Bislama, Crowley (1990, 203) claims that *finis* had become firmly established as a perfective marker between the 1880s and the mid of the 20th century, having been a productive pattern in the 1890s at the latest. We should assume that a similar basis for this pattern existed in the early forms of Tok Pisin at the time, given that up to the end of the sandalwood era "the tense-aspect-mood systems of all Melanesian Pidgin speakers were probably fairly similar" (1990, 201). We can also assume that *bin*, at that time, still served as a general past marker, closer semantically to the concept of a completed action than its later function as perfect and pluperfect. It is this closeness in function which could have brought it into competition with *pinis*, as the concept of FINISH frequently develops into either completive, perfect or past tense (Bybee et al. 1996, 51). So rather than two pastness markers, as Aitchison describes above, speakers of Tok Pisin would have had access to a general past marker and a completive marker slowly expanding its function to past marking as well. However, that competition never fully played out, due to the fact that there was simply no strong functional demand for simple

past marking. As described above, most sentences in Tok Pisin either leave temporal reference unmarked, or mark it through adverbials. Hence, *bin* as a simple past marker was as unnecessary as an expansion of *pinis* to that function. Rather, *bin* eventually came to be used as a marker of anteriority - in situations which required specific tense relations to be expressed, while *pinis* retained its status as a completive marker. The eventual lack of competition on a functional level - one expressing a tense relation and the other an aspect - further benefitted from a lack of competition on a syntactic level: since one of the markers was preverbal and the other postverbal or clause-final, neither replaced the other despite the relatively long time period since their inception. The eventual use of *pinis* with stative predicates would have further delineated it from *bin.*

It is difficult to say, of course, which would have 'won out' had past marking been in functional demand. There are arguments to be made for both *bin* as an established tense marker and *pinis* as being closer conceptually. There might also have been a link in mood type, given that *pinis* by virtue of its semantics is closely linked to the realis, while *bin*, in the hypothetical past, can also express irrealis moods. In any case, their lack of competition is exemplified in the fact that they can now occur together within one verbal construction - a pattern we are also about to observe for *wok long* and *i stap*. In Bislama, on the other hand, *bin* and *finis* are "semantically incompatible" (Crowley 1990, 205), since *bin* marks only the prior past, not also an anterior to nonpast as *pinis* does in Tok Pisin.

6.8.2 The continuous complex: *i stap* and *wok long*

Tok Pisin is capable of marking both continuous and progressive aspect, i.e. incomplete actions and actions in progress at a specific time, respectively. The marker for these aspects, *i stap* is another of Tok Pisin's grammatical items that may seem counter-intuitive to a speaker of English at first. Why would an adapted form of English *stop*, which implies the cessation of an action, indicate the exact opposite: a continuation of said action? It should come as no surprise at this point that once again, this form-function pairing is not a common one among English-lexified contact languages, as table 6.22 shows.

Language	Type of progressive marking	Marker
Early Sranan	progressive, habitual, current	*de*
Sranan	progressive, habitual, current	*e*

Saramaccan	progressive, habitual, current, future	*ta*
Nengee	progressive, habitual, current, future	*e)*
Creolese	progressive, habitual, future	*a*
Trinidad Endlish Creole	only progressive	*be + -ing*
Vincentian Creole	progressive, habitual, future	*a*
Jamaican	progressive, habitual, future	*de / a*
Belizean	progressive, habitual	*de*
San Andreas Creole English	only progressive	*de*
Nicaraguan Creole English	only progressive	*de*
Bahamian Creole	progressive, habitual, current, future	*-ing*
Gullah	progressive, habitual, current, future	*da*
African American English	only progressive	*-ing*
Krio	only progressive	*de*
Ghanaian Pidgin English	progressive, habitual, current	*dè*
Nigerian Pidgin	progressive, habitual	*dè*
Cameroon Pidgin English	progressive, habitual	*di*
Pichi	progressive, habitual, current, future	*dè*
Chinese Pidgin English	no overt progressive marker	-/-
Singlish	progressive, habitual	*-ing*
Tok Pisin	only progressive	*i stap / wok long*
Bislama	progressive, habitual, current	*i stap*
Norf'k	only progressive	*-en*
Kriol	variable only progressive	*-(a)bat / -ing*
Hawai'i Creole	only progressive	*-in*

Table 6.22: Progressive aspect systems in the 26 English-lexified languages in the APiCS data

Mühlhäusler (1985, 379f.) notes the following in regard to the continuous in

Tok Pisin:

> Predicates are either followed by *i stap* or preceded by *(i) stap* to indicate actions or states which are continuous , translating English progressive forms involving a form of to be and a verbal form ending in *-ing*. In the case of predicates containing transitive verbs, *i stap* may follow directly after the verb. [...] The use of *stap* or *i stap* implies no judgement about the length of time for which an action or state continues. If the speaker wants to indicate that an action continues for a long time, he/she can repeat either the main verb or *i stap* in postpredicative position[.] Alternative ways of expressing the progressive aspect are the use of the adverb *nau* sentence finally and the use of wok *long* followed by the predicate; *wok long* has gained considerable popularity in recent years through its use in radio broadcasts.

Both options are present in the Z'graggen corpus, as are sentences in which both strategies are employed simultaneously, as examples 127 through 132 show:

(127) Orait muruk i dai **i stap** em larim na ron i
Alright cassowary PM die PM PROG 3SG leave CONJ run PM
kam bek long ples.
come back PREP-D place

Alright, the cassowary is dying, it leaves and runs back to the place. (ZC 00036)

(128) Na ol i wokim haus nau, tupela i wokim haus **i**
CONJ 3PL PM build house now PRN PM build house PM
stap hia.
PROG here

And they build a house now, both are building a house here. (ZC 02518)

(129) Na sampela lain ol **i wok long** kisim taro nau.
CONJ DT family 3PL PM PROG PROG get taro now.

And some families are picking taro now. (ZC 00264)

(130) Em i kam bek na liklik brata bilong em sindaun
3SG PM come back CONJ small brother PREP-P 3SG sit-down
antap long diwai kalapulim na **i wok long** singautim
on-top PREP-L tree kalapulim CONJ PM PROG PROG call
si. (ZC 03092)
sea

He comes back and his small brother sits on top of a kalapulim tree and is calling out to the sea.

(131) Tupela i   kam na   sampela man i   **wok   long**   singsing
     PRN   PM come CONJ DT   man PM PROG PROG sing
     i   stap.
     PM PROG.

     The two come back and some man is singing. (ZC 00134)

(132) Na   ol i   **wok   long** lukluk **i   stap**, na   em i
     CONJ 3PL PM PROG PROG look   PM PROG CONJ 3SG PM
     kam sua long   ol.
     come shore PREP-D 3PL.

     And they are watching, and he comes ashore to them. (ZC 00371)

Meyerhoff (2013, 47) provides two examples of the *i stap* progressive for Bislama (note both HAB/PROG in second example):

(133) mifala   **i   stap** talem wan gel ia   hem i   bin bon
     1PL-EXCL AGR PROG tell   INDF girl DEF 3SG AGR ANT burn
     long faea
     LOC fire

     We're talking about this girl who was burnt in a fire.

(134) mi **stap**   hetem hem bifo   finis
     1SG PROG/HAB hate   3SG before COMPL

     I hated her already from before.

Solomon Islands Pijin, on the other hand, uses a different kind of progressive marker (Huebner and Horoi 1979, 91):

(135) Mere   ia   hem i   **gohed fo** widim gaden.
     Woman FOC 3SG PM PROG   PROG weed   garden

     That woman is/was/will be weeding the garden.

(136) Kupe hem i   **gohed fo** wakem haos.
     Kupe 3SG PM PROG   PROG build house

     Kupe is/was/will be building the house.

Once again, we see Tok Pisin and Bislama sharing a TMA marker, while Pijin deviates from the common pattern. This suggests that either Bislama and Tok Pisin innovated the same marker, or that it was present in some earlier stage of Melanesian Pidgin. Before going into the origin of the construction, however, *i stap*'s competition needs to be mentioned. In Tok Pisin, the progressive aspect can also be expressed by "the use of *wok long* followed by the predicate" (Wurm and Mühlhäusler 1985, 280). *Wok long* is certainly a highly

frequent construction in the Z'graggen corpus, with 18023 occurrences. Interestingly, sometimes speakers will employ both *i stap* and *wok long* in the same predicate:

(137) Na   pikinini ia   wok   long   krai, krai, krai **i   stap**   [...]
       CONJ child   FOC PROG PROG cry   cry   cry PM PROG [...]

And the child was crying, crying, crying [...]. (ZC 03127)

(138) Na   em tok, Ah yupela **wok   long**   giamanim mi   ah?
       CONJ 3SG say Ah PRN   PROG PROG lie   1SG ah

And he said, ah, you two are lying to me, eh? (ZC 02482)

Before taking a closer look at the gramamticalization of *i stap* and *wok long*, it is necessary to first distinguish closely between the progressive and continuous aspects. Both refer to an incomplete action or state in progress at a specific point in time. While the progressive aspect focuses on the action, the continuous aspect focuses on the state of the entity engaging in the action. Although they cannot always be distinguished in many languages including English without additional structural material - *She is learning German* is parsable both as continuous and progressive, while *She is learning German right now* would be more likely to be parsed as progressive rather than continuous - this is not necessarily the case in all other languages. The first two questions, then, are whether Tok Pisin distinguishes between the continuous and progressive aspects and if so, whether *stap* and *wok long* can be used to express both the continuous and progressive aspect or just one of the two. While *i stap* allows for both a continuous and a progressive reading, *wok long* is limited to progressive parsing of the predicate it occurs with. This is exemplified in the excerpts from the Z'graggen corpus above. In example 128, which uses *i stap*, the process of building a house suggests an ongoing, continuous process that, due to its duration, shifts the focus onto the subjects and therefore the continuous reading of the subjects being in the state of building a house. The cassowary dying in example 127, which also makes use of *i stap*, on the other hand, is indicative of a process very much tied to that specific period of time, as are the watching people in 132. As such, they benefit a progressive reading in which the process is focused. Similarly, the families picking taro in example 129 for *wok long* suggests an immediate, focused process, as does the brother calling out to sea in example 130, and the man singing in 131. However, the examples are not one hundred percent definitive, and could theoretically allow for both readings.

Still, the notion of *wok long* being limited to the progressive is also reflected in its most common verbal collocates: they are, in order, *kaikai* 'to eat', *krai* 'to cry', *lukluk* 'to see', *painim* 'to search', *kisim* 'to get', *singsing* 'to sing', *mekim* 'to make/build', *katim* 'to cut', *lukautim* 'to care for', *sutim* 'to shoot', *toktok* 'to talk (to)', *singaut* 'to shout', *paitim* 'to fight', *bihainim* 'to follow' and *tingting* 'to think'.

A closer look at *mekim* reveals the following tendencies: *wok long mekim* is mostly used in constructions such as *wok long mekim paia*, *wok long mekim nois*, *wok long mekim kaikai* or *wok long mekim hul*. In other words, processes which are fairly limited in time.

The additional function of *i stap* as a continuous becomes relevant when examining the possible grammaticalization paths the two forms may have taken. The continuous[89], as Heine and Kuteva (2002, 330) refer to it, is among the functions with the highest amount of possible source concepts in their work. They include COME, COMITATIVE, DO, EXIST, GO, IN (SPATIAL), KEEP, LIE, LIVE, LOCATIVE, LOCATIVE COPULA, SIT and STAND. I will argue that the two forms *stap* and *wok long* have grammaticalized from two different source paths and it is due to this different origin that one is used as both the continuous and progressive, while the other is limited to merely the progressive.

The only grammaticalization path for STOP is into the prohibitive (2002, 325). There is no salient semantic property of 'stop' that would directly indicate a continuous/progressive action. At the same time, it seems unlikely for the form *stap* to derive from any other etymon than *stop*, especially given the fact that the necessary phonological adaptation from /ɒ/ to /a/ not only requires a single step, but is also paralleled in similar developments such as /a/ in *tang* (from English ʌ in *tongue*) and the intransitive form *stapim* 'to stop sth.' exhibits the same vowel quality. How then do we get from English *stop* to Tok Pisin's *stap*?

I see two possible grammaticalization paths for this development, one mainly semantic in nature, while the other relies on syntactic reanalysis. The former is based on the secondary meaning *stap* acquired as a lexical verb before its grammaticalization to an aspectual marker. Hall's (1943a, 118) vocabulary section lists the meanings and functions of *stap* as nominal 'cessation' and as an intransitive verb meaning 'cease, stop; stay, remain; continue, be (continually, always); exist, live'. Mihalic (1957, 137) lists two distinct sets of verbal

---

[89]Unfortunately, Heine and Kuteva do not differentiate between the continuous and progressive aspects (and, additionally, also subsume the durative aspect under the umbrella of the continuous).

meanings for *stap*, namely 'to stop, to stay, to remain' and 'to be, to exist, to live, to be present, to be located', along with the function of "express[ing] the English durative or continued action or state...whenever an English verb form ends in 'ing'". It is in this secondary existential meaning that we find the grammaticalization path to the progressive - note that one of the originating concepts of the progressive as described by Heine and Kuteva above is EX-IST. This path would require a semantic expansion of the original borrowed meaning of *stop* 'to cease' to 'remain, exist, be', as described in the dictionary entries above. This requires three steps, the first being from 'stop' to 'stay' or 'remain', the second from 'stay' to a locative use, and the last from the locative use to an existential. [90] The semantic link relevant to the first step should be obvious: if one stops, one stays in place, or remains. Consider an English sentence such as *We made a stop at the local diner* or *We stopped by for a quick visit.* In such sentences, it is clear that *stop* does not (primarily) indicate a cessation of movement or activity, but the fact that the speaker remained at the aforementioned place, and therefore was located there for a while. For the next step, the place in which one remains is the place one is located. In a further step, this meaning would have been extended to the existential 'be', as observed in examples 137 and 138 above. Once again, individual steps of such a development are well attested, with LOCATIVE being one of the possible sources of the EXIST concept in Heine and Kuteva (2002, 330), while Siegel (2000, 214) documents the shift from *stop* to *stay* for the locative in Hawai'i Creole. The grammaticalization of *stap* then proceeds further from the existential to the continuous and to the progressive. It can therefore be summed up as follows:

$$\text{stop} \Rightarrow \text{English lexical item}$$
$$\Downarrow$$
$$\textit{stop/stap} \Rightarrow \text{Tok Pisin lexical item 'to cease'}$$
$$\Downarrow$$
$$\textit{stap} \Rightarrow \text{TP 'cease', TP 'stay, remain'}$$
$$\Downarrow$$
$$\textit{stap} \Rightarrow \text{TP 'cease', TP 'stay, remain', existential}$$
$$\Downarrow$$
$$\textit{stap} \Rightarrow \text{TP 'cease', TP 'stay, remain', existential, continuous}$$
$$\Downarrow$$
$$\textit{stap} \Rightarrow \text{TP 'cease', TP 'stay, remain', existential, continuous, progressive}$$

---
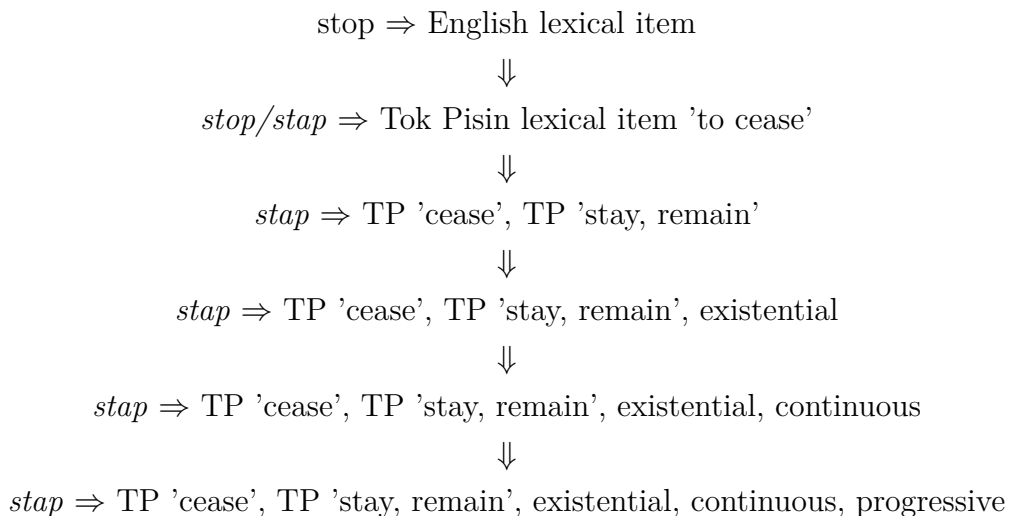
[90]Alternatively, the path could also have lead from 'stay' directly to 'exist', although I believe that would have required additional motivation.

222

Such a grammaticalization process is by no means singular to Tok Pisin. Winford (2012, 445) cites several examples and points out the following:

> Similar developments have been reported for other creoles. In the Pacific, Hawai'i Creole employs pre-verbal *stei* 'be at, stay' (< *stay*) as a marker of Progressive aspect. Tok Pisin and Bislama, respectively, employ post-verbal *i stap* and preverbal *stap* 'stay, be at' in the same function. Siegel (2000 , p. 219f.) argues that the progressive function of *stei* is modeled on the similar function performed by Portugese *esta*. But it is also quite possible that it is modeled on the use of Portugese *ficar* 'stay' as a progressive marker, as Sanchez (2006, p. 291) suggested. As far as the Melanesian Pidgin varieties are concerned, the use of *stap* as a marker of Progressive seems to be modeled on the use of existential verbs as markers of progressive aspect in many Oceanic languages (see Siegel, 2008, p. 188f. for fuller discussion).

For *wok* on its own, on the other hand, Hall records no aspectual function, merely nominal 'work, garden' and the transitive verb 'make'. For *wok*, Mihalic records 'to work at, to be busy about, to be engaged in', citing the example *Em i-wok long painim kaikai.* 'He is busy looking for food' (Mihalic 1957, 137).

Wok long could conceivably be a combination of two of the grammaticalization paths mentioned above for the continuous, *wok* representing DO and *long* representing the spatial aspect present in IN (SPATIAL.) As for the first concept, it seems fairly obvious that verbs with the meaning of 'do' are already semantically bleached to some extent and, due to their status as an auxiliary verb and consequent reliance on lexical verbs, make a prime candidate for grammaticalization. In regards to the spatial concept, Heine and Kuteva (2002, 179) describe it as "an instance of a more general process whereby grammatical aspect functions are conceptualized and expressed in terms of locative concepts", as is the case in a possibly emerging progressive aspect in German *am + V + sein*. English also has a similar construction in *work on (doing) something*, with *on* providing a similar function to *long* in *wok long* in sentences such as *I'm working on getting the garden winter-ready*. Note how the action itself takes the syntactic form of a prepositional complement, which, if a similar process underlies the construction in Tok Pisin, would explain the prepositional marker *long*. It is conceivable that *long* is playing a similar role in *wok long* as it has in the grammaticalization of *wantaim* (see chapter 6.2 above). Note, however, that unlike *wantaim (long)*, *wok long* seems to preserve its two constituents. There are only very few occurrences in which *wok* is directly followed by a verb: 51 for *kaikai*, 17 for *mekim* and 13 for *kisim*

being the most frequent ones. The first of these seems to be an idiom meaning 'to work for board and keep' and is therefore even less relevant.

Whether or not *long* has played an equal or a secondary role to *wok* in the construction, it seems evident that unlike *i stap*, *wok long* grammaticalized from the concept of DO. As I have hinted at above, this partly explains the function of *i stap* as being able to mark both progressive and continuous aspects, whereas *wok long* is mostly limited to the progressive aspect: *wok long* is both the more recently grammaticalized and therefore newer form (Wurm and Mühlhäusler 1985, 129) and the one more closely linked to the action concept of DO. As such, it is more likely to focus solely on denoting actions rather than states. *I stap*, being the older form coming from the existential grammaticalization path, might have denoted only a continuous aspect at first, but has had time to expand its function to the progressive as well. Similar to Haspelmath's semantic maps I have re-drawn for *wantaim* above (see 6.2), we might therefore establish the following semantic map or grammaticalization path 6.8 for the continuous and progressive aspects in Tok Pisin:

```
progressive  ──  continuous  ──  existential
     |                               |
   'do'                           locative
                                     |
                              'stay, remain'
                                     |
                                  'stop'
```
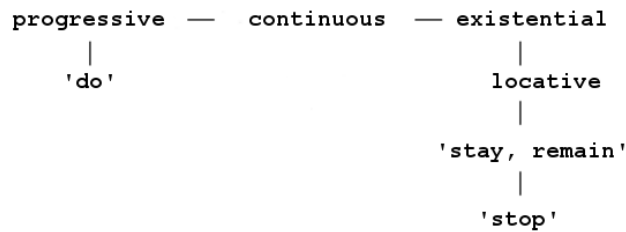
Figure 6.8: Grammaticalization path for the progressive

As was the case with *na* and *wantaim*, both grammaticalization paths in question are equally valid, but the distance from the starting points DO and STOP in both a temporal and a conceptual sense, are not equal. STOP, or *stap*, has had a head start, as described above. DO, or *wok*, is only now catching up, and while it is perfectly plausible for it to compete with *stap* for the continuous aspect eventually, that does not seem to be the case in the Z'graggen data yet. Table 6.23 below illustrates the distribution of *i stap* and *wok long* for 19 different verbs of interest in the Z'graggen corpus:

Verb	*wok long*	*i stap*	Ratio
singautim (call, demand)	60	0	-
askim (ask)	149	6	24.83
bihainim (follow, obey)	299	14	21.36
ron (run)	66	7	9.42

katim (cut)	409	53	7.72
singaut (call, demand)	319	104	3.07
tingting (think)	287	111	2.59
winim (win)	56	22	2.54
kukim (cook)	209	111	1.88
kaikai (eat)	979	1190	0.83
lukluk (see, watch)	686	902	0.76
krai (cry)	465	832	0.56
singsing (sing, dance)	535	1180	0.46
wokim (do, make)	214	479	0.45
pundaun (fall down)	31	115	0.27
kamdaun (come down)	36	145	0.25
putim (put)	147	1130	0.13
slip (sleep)	118	1732	0.07
hait (hide)	49	681	0.07

Table 6.23: Distribution of *wok long* and *i stap* across 19 verbs in the ZC

There are two observations to be made from this table[91]. The first is about the extreme ends. For verbs such as *slip* or *hait*, *i stap* is strongly preferred. This is very likely to be because of the fact that the action of sleeping is less likely to be focused on than the state of the sleeper, i.e. the state of the subject. In other words, the information is more likely to be that a person is asleep than the action of sleeping being carried out. The same is true for hiding. The fact that *putim* also preferably occurs with *i stap* is harder to explain in this regard. Why would the focus be on the state of the person putting something down? Part of the answer is that putting a single object down is too brief an action for the progressive to become necessary. Once multiple objects are involved, the focus is once more likely to be on the subjects involved.

Let us consider an English equivalent of *they're putting up a fence.* There is an argument to be made that the longer the duration of an action, the more likely it is to be backgrounded against the subjects involved, and the more likely it is to occur in the continuous than in the progressive. An example in English would be the progressive versus the continuous reading of *I'm learning Spanish* (right now, progressive) and *I'm learning Spanish* (continuous). This can also be observed in the other end of the table, with the verbs occurring

---

[91]The p-value for the difference between all the verbs described in the following paragraph is significant at p below 0.05.

predominantly with *wok long*. *Askim*, i.e., to ask, is usually a singular occurrence. It is unlikely - though certainly possible - for someone to be in the state of asking. The same is true for *bihainim* and *ron*, especially since in the data, *ron* does not occur in the sense of running as exercising, which makes it more likely to occur in the continuous. *Katim* as well, usually describes a singular occurrence, as does *singaut*. Secondly for verbs readily parsable as either progressive or continuous, such as *kaikai, lukluk, krai, singsing* and *wokim*, we see a more even distribution than for the ones above, but a distribution that is still slanted in favour of *i stap* as the older, more established, more entrenched aspectual construction. The pattern becomes clearer in the mosaic plot below:
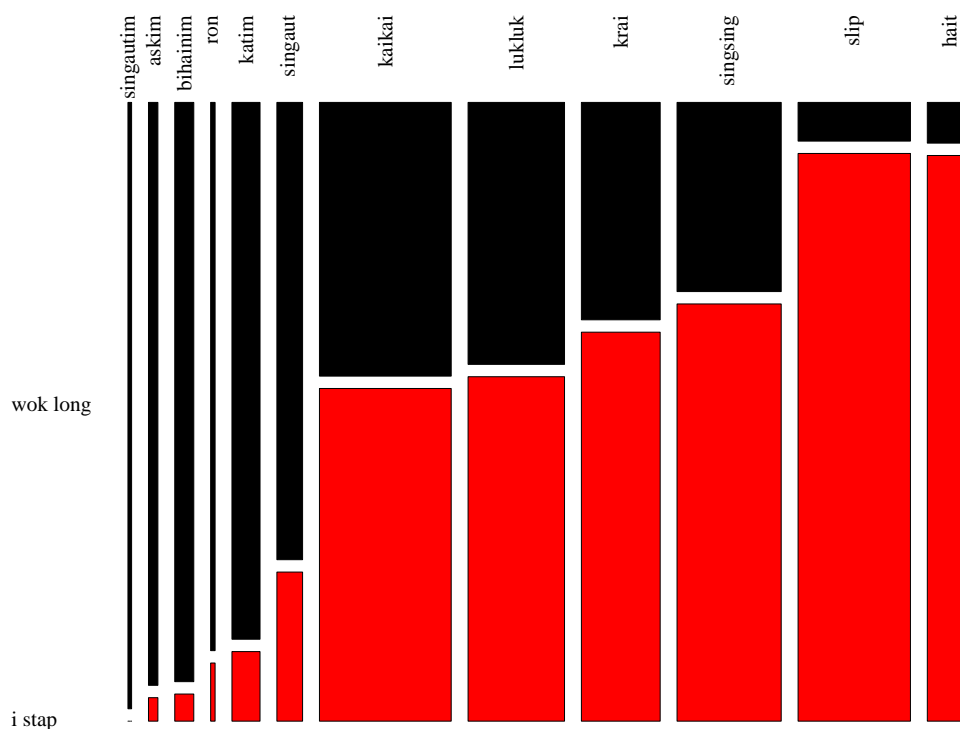
Figure 6.9: Distribution of *i stap* and *wok long* with key verbs

To further validate the difference in function between *wok long* and *i stap*, a qualitative analysis of a sample of 500 occurrences of *wok long* was carried out. Among these 500 occurrences, 431 were of a progressive function, while only 2 could be parsed as primarily continuous. Interestingly, in 34 cases, verbs occured with both *wok long* and *i stap*, which can be taken as further proof that they serve two distinct and complementary functions as a progressive and continuous marker. Consider, for instance, the examples below:

(139) Na   taim em wok   long kaikai i   stap, bai yu
      CONJ when 3SG PROG PROG eat   CONT CONT FUT 2SG
      rausim   olgeta purpur bilong   yu.
      throw-out all   flower   PREP-P 2SG [...]

And when he is eating, you will throw out all of your flowers. (ZC 01018)

(140) Na   em wok   long slip i   stap   arere   long
      CONJ 3SG PROG PROG sleep CONT CONT PREP-L PREP
      dispela hap wara, em i   no   inap long   muv.
      DT   place water 3SG PM NEG ABL PREP move

And he was sleeping next to the water, he was unable to move. (ZC 02537)

In both of these cases, it could be argued that the progressive aspect through *wok long* could be invoked to convey that the action was unfinished, and the continuous aspect invoked to convey that the person in question is in a state of eating or sleeping (i.e., unable to observe the action described in the last part of the first sentence and unable to move as described in the second example).

As usual, we also have to consider possible influence from the Austronesian and Oceanic substrates. As Siegel (2008, 190) points out, "there is no progressive construction in Tolai and related languages which is parallel to that of Tok Pisin". Instead, he reports, referring to Mosel's work, Tolai makes use of reduplication strategies for the progressive and other aspects (Siegel 2008, 190), (Mosel 1980, 102), (Mosel 1984, 97). He further quotes Crowley (1990, 218) in saying that "serial constructions with locative/existential verbs in Tolai and related languages may have predisposed speakers to the V *i stap* construction rather than *stap* V".

In regard to the progressive and continuous aspects in Tok Pisin, then, we can conclude with three observations. First, the fact that *stap* has proceeded on an unusual, but in no way a unique grammaticalization path. Second, that *wok long* has grammaticalized later and on a different path than *stap*, and the two constructions are not (yet) equal in their functional expanse. Third, that substrate reinforcement is likely to have played a larger role in the developmental step of *stap* from existential to aspect marker. We will see some of these patterns and observations emerge once again in the remaining TAM markers below.

6.8.3 The inceptive complex: *kirap* and *kamap*

The inceptive complex consists of the ingressive and inchoative aspects. These are frequently summarized under change-of-state (Heine and Kuteva 2002, 18), though some languages mark a distinction. Defined narrowly, it is the inchoative aspect that marks a change of state, or process, while the ingressive aspect marks the start of an action. In English, an example of the inchoative would be *The flower turned red*, whereas the ingressive could be exemplified by *They started working on the house*. Note that as a verbal aspect, however, they are neither grammatically distinct in English, nor are they grammaticalized as aspect markers. Inchoative sentences such as *The floor started burning* and ingressive sentences such as *He started running* exhibit no structural difference.

Tok Pisin has three possibilities of expressing change-of-state in a wider sense. They are *kirap*, *kamap*, and *stat (long)*. In general, all of these can be used to express both the inchoative (i.e., change-of-state in a narrower sense) and the ingressive. However, when it comes to doing so in form of an aspectual marker, there seem to be strong preferences among speakers of Tok Pisin. The examples below show each expressing an ingressive action as either an aspectual marker or a serial verb construction (a distinction that will be of relevance below):

(141) Orait Manub wantaim Kulbob i   **kirap** lusim dispela asples
       Alright Manub PREP-C Kulbob PM ING   leave DT   home
       bilong   ol   nau.
       PREP-P 3PL now

       Alright Manub and Kulbob start to leave this home of theirs now. (ZC 00185)

(142) Ol man **kamap** sutim dispela Tumiange nau.
       PL man ING   shoot DT   Tumiange now

       The men started shooting this Tumiange then. (ZC 03575)

(143) Na   ol i   go, ol i   **stat long** katim dispela diwai.
       CONJ 3PL PM go, 3PM PM ING   ING   cut   DT   tree

       And they went, they started cutting that tree. (ZC 00913)

Of these three, *stat* is of the least interest to the present analysis for two reasons. First, it is by far the least commonly employed of the three options. Overall, *stat*, in all of its functions, occurs only 1,400 times in the Z'graggen corpus. *Kirap* and *kamap*, on the other hand, occur 19,852 and 45,759 times, respectively. Table 6.24 below shows the number of occurrences for each of the three options before eighteen verbs:

Verb	*kirap*	*kamap*	*stat (long)*	Ratio kirap/kamap
kisim (get)	968	104	31	0.107
tokim (talk, speak)	753	20	6	0.027
bihainim (follow, obey)	441	32	6	0.073
tok (talk, say)	351	22	1	0.063
ronowe (run away)	333	6	1	0.018
lusim (lose, forget)	264	10	6	0.038
wokabaut (walk around)	244	16	18	0.066
kilim (beat, kill)	229	54	11	0.235
karim (carry, endure)	216	47	7	0.218
katim (cut)	163	18	65	0.110
sutim (shoot)	136	21	15	0.154
askim (ask)	129	17	3	0.132
mekim (make)	113	69	19	0.611
lukim (see)	111	132	0	1.189
brukim (break)	108	22	11	0.204
paitim (beat, hit)	98	12	6	0.122
tanim (turn into/around)	97	19	1	0.299
wokim (make, do)	96	35	47	0.365
Total	4,850	656	254	-

Table 6.24: Distribution of *kamap* and *kirap* across 18 verbs in the ZC

As can be seen, *kirap* is by far the most preferred option for the ingressive in the majority of verbs. *Stat*, on the other hand, is statistically irrelevant for many of them. In addition, it is clearly adapted from English *start*, which makes it of lesser interest in that it is not an innovation of Tok Pisin. It will therefore be disregarded for the rest of this chapter.

Both *kirap* and *kamap* in the function of ingressive markers, however, are, if we are to assume that they indeed are aspectual markers. In this regard, Mühlhäusler writes:

> In opposition to the aspect markers treated so far, *kirap* shares all the properties of real verbs, and its occurrence as an aspect marker can be regarded, syntactically, as an instance of verbal concatenation. *Kirap* directly preceding other verbs indicates actions which are [begun][92] at the time referred to in the sentence in which it occurs[.]

---

[92] The original text says "being" at this point, which I consider to be a typo.
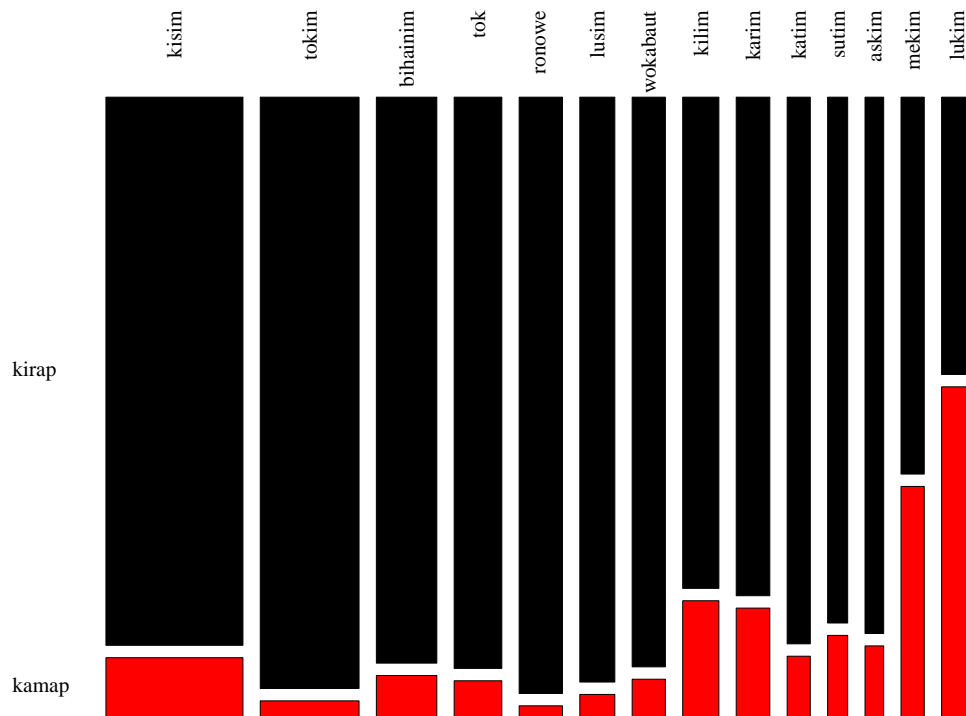
Figure 6.10: Distribution of *kamap* and *kirap* across 14 verbs in the ZC

What are the factors which speak against *kirap* (and *kamap*, incidentally) being considered aspectual markers rather than just verbal concatenation? If we apply Lehmann's (2015) parameters of grammaticalization, three are not present: there is no phonological or morphological reduction between *kirap* and *kamap* as full verbs and them being used as ingressive markers. Note that this has, apparently, happened in Bislama. *Kirap*, "often in the shortened forms *kira*, or even *kra* -is increasingly used in colloquial Bislama to indicate that an action happens after something else in a sequence of happenings, especially when somebody is telling a story" (Crowley 2004, 172):[93]

(144) Olfala   we   i   stanap long   stoa i   **kirap** i   tok
       Old-man REL PM stand   PREP-L store PM SEQ   PM tell-off
       long   ol pikinini we   oli   mekem noes.
       PREP PL child   REL 3PL make   noise
       The old man standing at the shop told off the noisy children.

Secondly, they are not obligatory by any means. Third, they freely combine with any other number of structures in their use as full verbs. At best, then, they are currently - or as of the time of the Z'graggen corpus - candidates for

---

[93]Translation by Crowley, glossing mine.

future grammaticalization. However, this does not indicate by any means that their origin is not of interest here. A hint towards the respective origins lies in their use not in an ingressive, but as a full verb with a inchoative or change-of-state meaning, akin to English *become*. As with the ingressive function, both *kirap* and *kamap* can be used in this way:

(145) Na   man ia   i   laik kaikai nau paia i   **kamap** nau, paia
       CONJ man FOC PM want eat   now fire PM start   now fire
       i   lait   na   i   **kirap** bikpela.
       PM light-up CONJ PM INCH   big
       And the man wanted to eat now, the fire started, the fire lit up and became big. (ZC 02191)

(146) Flawa bilong   ol i   **kamap** retpela ia.
       Flower PREP-P 3PL PM INCH   red   FOC
       Their flower turned red. (ZC 03407)

However, it is *kamap* which occurs more frequently in this function. *Kamap *pela* occurs 2,132 times in the Z'graggen corpus, whereas *kirap *pela* occurs only 158 times. Therefore, *kamap* seems to have a stronger link to the change-of-state concept, whereas *kirap* seems to be the preferred option for the ingressive. This is also reflected in constructions such as *plaua kamap*, which translates as 'the flower blooms'. How could such a tendency be explained? It is by the inherent semantics of the verbs these markers derived from.

The origin of Tok Pisin *kirap* lies in the English *get up* (Mosel 1980, 17). It retains this meaning as a full verb both in Tok Pisin and in Bislama in sentences such as *Taim mi lusim slip na mi kirap* 'When I wake up and get up' (Verhaar 1995, 133) and, *Yu kirap long stul* 'Get up from the stool' (Crowley 2004, 172) respectively. It is not difficult to construct a semantic link between getting up and starting to do something: by definition, one was doing nothing (leaving aside dreaming and sleeping) before one gets up. Which is why the event that occurs next has to start, rather than having been ongoing. Hence even when *kirap* is used in its original sense of getting up, it frequently describes the first in a series of events rather than an event of its own. Unsurprisingly, *na* is its most frequent collocate, with 2590 co-occurrences. The construction *kirap na* in turn is frequently followed by *tok, tokim, kisim, kilim, wokobaut, askim, bihainim* and similar verbs.

Kamap, on the other hand, derives from English *come up*. Mihalic (1957, 53) records its meanings as 'to come up, to appear, to rise, to grow', 'to begin' and 'to come to, to arrive at', 'to reach'. Heine and Kuteva (2002, 328) list COME

TO, GET and GO as source concepts for the grammaticalization of constructions expressing change-of-state, further noting that the path from COME TO "appears to be particularly common in pidgin and creole languages" (2002, 74). While they make no distinction between the ingressive and inchoative, treating both as change-of-state *kamap*, at least, was first, or at least, primarily used to express an inchoative function, i.e. change-of-state in the narrow definition, from which it expanded to an ingressive function. In this non-aspectual inchoative function, Verhaar (1995, 84f.) calls it "copulative *kamap*", also noting that "all clauses with copulative *kamap* are characterizing; not identifying". *Kamap*, due to its origin, still exhibits a semantic link to the notion of both a spatial distance and a directed movement through that distance in its function as an ingressive marker. Consider the verbs in table 6.24 above with which *kamap* occurs uncharacteristically rarely: among them are *lusim* 'leave' and *ronowe*, which both imply a movement away from a point, whereas *kamap* implies a movement towards that point. Furthermore, many of the verbs it appears with more frequently - such as *kilim, karim, mekim, brukim, tanim* and *wokim* necessitate a physical proximity first established by coming towards something. It is this concept which aligns well with the process nature of the inchoative[94].

In competing with *kirap* for the ingressive function of the inceptive complex, then, *kamap* is hindered by such a limit, which partly explains its lower frequency as an aspectual marker despite its higher number of overall occurrences as a lexeme.

6.8.4 The abilitative-permissive complex: *save, inap* and *ken*

In modern Tok Pisin, *inap* has several distinct functions. Verhaar (1995, 137f.) sums them up as follows:

> *Inap* as a modal auxiliary must be distinguished from *inap* as a modifier, meaning 'sufficient', and it may then also function as a stand-alone predicate, as in *Em inap*, which may mean 'That's enough' or 'That is sufficient' [...]. Also, *inap* or *inap long* may mean 'as far as' or 'until', as in *inap long banis* 'as far as the wall' or in *inap long belo kaikai* 'until noon' [...]. Finally, *inap long* may function as a conjunction, to introduce a clause within a complex sentence [...].

---

[94]Incidentally, this is a further indication for *kamap* not being heavily grammaticalized yet: it lacks semantic bleaching.

None of these three are auxiliaries, although there are gray areas between inap as the modifier 'sufficient, enough' and the auxiliary verb 'to be able to, to be capable of' [...].

All of these functions can be observed in the Z'graggen corpus, as the following examples will show. The first two illustrate its function as a modal auxiliary, i.e. the abilitative use, in which it marks the ability of the subject to carry out an action:

(147) Mitupela   tete   bai   i   no   **inap** go bek long   ples
      1PL-DUAL today FUT PM NEG ABL go back PREP-D place
      long   lukim papa mama bilong   mitupela   nau.
      PREP-F see   father mother PREP-P 1PL-DUAL now
      Now we will not be able to go back to the place to see our father and mother today. (ZC 03030)

(148) Tasol pren bilong   em tokim em, Nogat, mi   no   **inap** givim
      CONJ friend PREP-P 3SG tell   3SG NEG   1SG NEG ABL give
      yu.
      2SG
      But his friend told him, No, I cannot give it to you. (ZC 03598)

The next two show *inap* as a standalone predicate:

(149) Na   em   tok, **Em inap**.
      CONJ 3SG say 3SG   enough
      And he said, that's enough. (ZC 01647)

(150) Ok   Father **em inap**.
      Okay Father 3SG   enough
      Okay Father, that's enough. (ZC 00690)

The following examples show *inap* as a modifier:

(151) Bai mipela   i   gat **inap** kaikai mipela   kaikai long
      FUT 1PL-EXCL PM have enough food   1PL-INCL food   PREP
      olgeta femili bilong   mipela.
      DT   family PREP-P 1PL-INCL
      We will have enough food we eat for our entire family. (ZC 004279)

(152) Ol i   traim best bilong   ol, i   no   **inap**, meri   ia
      3PL PM try   best PREP-P 3PL PM NEG enough woman FOC
      i   winim ol.
      PM beat   3PL

They tried their best, it was not enough, the woman beat them. (ZC 01253)

Examples 153 and 154 illustrate the function of *inap* as a preposition meaning 'as far as':

(153) I   go, i   go **inap long** ples   we   dispela talis i   stap
       PM go PM go as-far-as PREP place REL   DT   okari-tree PM
       [...]
       be   [...]

       He went as far as the place where that okrai tree is [...] (ZC 01030)

(154) Yes Luthren i   kam **inap   long** Malmal tasol, na   nau
       Yes Lutheran PM come as-far-as PREP Malmal only CONJ now
       mipela   olgeta i   kam long Katolik na   lusim Luthren.
       1PL-INCL PRN PM come PREP Catholic CONJ leave Lutheran.

       Yes, the Lutherans came only as far as Malmal, and now all of us
       became Catholic and left the Lutheran Church. (ZC 00173)

Examples 155 and 156, on the other hand, show the use of *inap* as a temporal conjunction:

(155) Em i   stap hia ating **inap long** taim   ol German man ol
       3SG PM be   here maybe CONJ CONJ CONJ PL German man 3PL
       i   kamap hia [...]
       PM come   here [...]

       He was here maybe until the German men came here [...] (ZC 00799)

(156) Em nau em katim rausim mekim hul **inap long** em i   go
       3SG now 3SG cut   remove make   hole CONJ CONJ 3SG PM go
       insait.
       inside

       Now, he cut and removed and made a hole until he went inside. (ZC
       03029)

The function of the abilitative can also, however, be expressed by *save*:

(157) Na   em **save** ronowe   wantaim dispela pikinini.
       CONJ 3SG ABL run-away PREP-C DT   child

       And he was able to run away with this child. (ZC 01979)

(158) Na   mama tok, Yes mi **save** lukim.
       CONJ mother say Yes 1SG ABL see

       And the mother said, Yes I can see it. (ZC 00965)

In Bislama, the abilitative, like the permissive (i.e., the grammatical mood indicating that an actions is permitted by the speaker or another person) and habitual (i.e., the grammatical aspect indicating an action occurs regularly or repeatedly) are expressed by *save* (Crowley 2004, 99) [95]:

(159) Mi **save** toktok Franis be i no turnas.
1SG ABL speak French CONJ PM NEG much

I can speak French but not too much.

(160) Yu **save** dring hamos sel kava?
2SG ABL drink how-much shell kava

How many shells of kava can you drink?

For Solomon Islands Pijin, Huebner (1979, 116) suggests that "*save* covers a wide range of meanings", including "physical ability , competence or habit and thus can be translated as either 'can' , 'know how to' , or 'be accustomed to'". He gives the following example, among others:

(161) Olketa i **save** fiksim motoka.
PRN PM ABL fix car

They are physically able to fix/know how to fix/are accustomed to fixing cars.

However, he also mentions two other possible forms with similar function: *fitim* and *inaf*, noting that the differences between them lie in the fact that while *save* can cover all of the meanings above, "*fitim* conveys the idea of competence, either physical, mental or emotional", while "*inaf* indicates skill at performing an action" (1979, 117). In the negated abilitative, there are even more possibilities, with *kanduit* 'cannot' competing with *no save, no fitim* and *no inaf*.

(162) Gele ia **inaf** fo lanem langwis.
Girl DT ABL PREP learn languages

The girl is skilled at learning languages. (Huebner and Horoi 1979, 116)

(163) Robert hem **inaf** fo fising.
Robert 3SG ABL PREP fishing

Robert is skilled at fishing. (Huebner and Horoi 1979, 116)

---

[95]Translation by Crowley, glossing mine.

Like Solomon Islands Pikin, Tok Pisin also features competing forms for the abilitative in *save*, *ken* and *inap*. Like the competing forms I have previously discussed above, these are not functionally identical, either. We shall first focus on the difference between *inap* and *ken*. Unlike *fitim* and *inaf* in Solomon Islands Pijin, the original functional difference between *inap* and *ken* in Tok Pisin lies not in competence versus skill, but in permission versus ability. While *inap* denotes the physical capability - that is, the abilitative in the traditional sense - *ken*, adopted from English *can*, mainly denotes the permission to perform an action (cf. Muysken 2008, 195f.), as seen in examples 164 and 165 below:

(164) Yu **no ken** i   go, yu mas i   stap.
       2SG NEG   PERM PM go 2SG must PM   stay

       You cannot go, you must stay. (ZC 01013)

(165) Na   tupela tok, Em i   orait yupela **ken**   kaikai, mitupela
       CONJ PRN   say 3SG PM alright 2PL   PERM eat   1PL-EXCL
       i   nogat tok.
       PM NEG talk

       And the two said, it's Alright, you can eat, we don't talk. (ZC 00698)

Both the abilitative and the permissive that the inference for the event the following verb describes is [+wanted] or [+desired], as opposed to the deontic inference [-wanted] that must carries (Brown 2008, 1015). Where they differ is whether the realization of said desire lies within the power of the speaker, which is the case for the abilitative, or with an external force, which is the case for the permissive. Languages such as English and German generally provide different grammatical means for expressing each of these moods (English *can* and *may*, German *dürfen* and *können*), although some overlap can occur: consider English requests such as *Can I have some water?* which, of course, do not refer to the speaker's ability to receive more water, but to an external force's permission to do so.

The functional difference described here has not been stable throughout the development of Tok Pisin. Hall's (1943a, 101) vocabulary section has four entries under *inap* (or rather, inɛf, as he spells it): adverbial 'enough, sufficiently', an intransitive verb 'be enough; suit, fit' and a minor clause 'that's enough of that' (presumably in combination with *em*).

Mihalic's dictionary lists four subentries under *inap*, which, he claims, derives from English. The first corresponds to both the functions of modifier and stand-alone predicate above, listing the meaning as "enough, sufficient" (Mihalic 1957, 49). The second and third are verbal in nature, with the recorded

meanings being "to fit, to be the right size, to be the right age" and "to be suited for, to be fit for, suitable, capable, able", respectively. Finally, the last entry points to the prepositional function, with the meanings recorded being "until, tull, up to, about". The third entry, which includes the abilitative, is of course of most interest to the point at hand. Mihalic seems not to have considered it a central function of *inap*, with the respective example "*em i-no inap long yu*" translated as "he is not good enough for you" (1957, 49). There is also no indication whether *inap* has a permissive function as well. The entry for *ken*, on the other hand, lists "can, be able, be possible, may, be willing to" as meanings of the word (Mihalic 1957, 58). In other words, it includes both the permissive and abilitative function.

More recently, Smith and Siegel (2013, 220) report that out of the two, *inap* is the one that can serve both functions, as it "can also be used in a permissive sense". The example they give is reproduced below[96]:

(166) **Inap** mi   baim dispela trausel long   yupela?
       ABL 1SG buy   DT   turtle   PREP 2PL
       Can I buy this turtle from you?

The expansion of *inap* to the permissive is a more recent development. In the Z'graggen corpus, there are still traces of a functional division between the two, which is evident, for instance, in their most common verbal collocates, summarized in table 6.25 below:

ken		inap	
stori (tell)	684	*lukim*	281
lukim (see)	679	*kisim*	156
kisim (get)	375	*kilim*	132
kaikai (eat)	374	*kaikai*	113
tok (talk, say)	251	*save*	83
harim (hear, listen)	195	*painim*	77
kilim (beat, kill)	179	*kam*	76
mekim (make)	168	*mekim*	53
kam (come)	78	*karim*	51
stap (stay, be)	62	*wokobaut*	46

Table 6.25: Most common 2R verbal collocates of *ken* and *inap*

---

[96]Translation theirs, glossing adjusted to reflect use within this study by me.
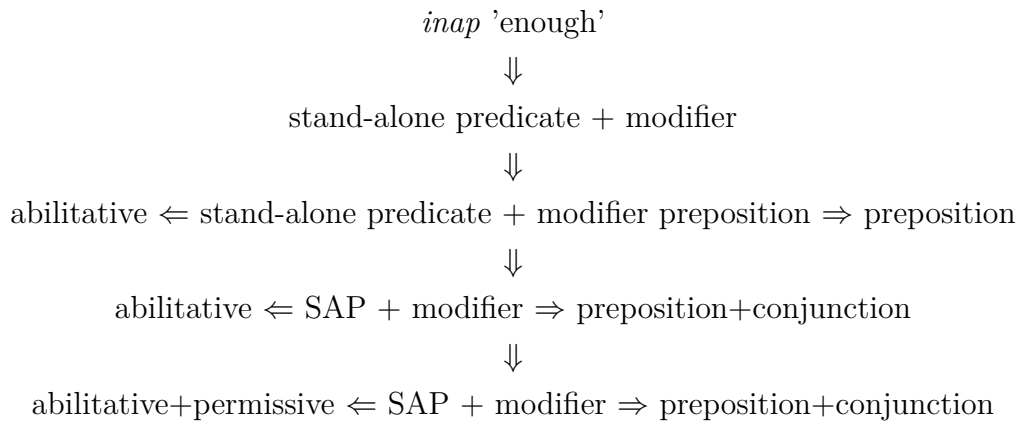
Stori 'to tell a story', for instance, occurs predominantly with *ken* at 684 occurrences, while it only occurs with *inap* 38 times. Naturally, one's physical ability to tell a story is much less often in question than one's permission to do so. *Lukim*, a verb that is bound to occur in the physical ability context more often than in the permissive, on the other hand, occurs predominantly with *ken* or *noken* (679 times), but also occurs with *inap* (281 times). Notably, it occurs with *inap* almost exclusively when negated: of the 281 occurrences, 270 are negated, while of the 679 occurrences with *ken*, only 272 are negated. This is a general trend for *inap*. Out of the 12,591 occurrences of *inap* in general, 6,474 are preceded by *no*. However, of the 588 times *inap* occurs followed by the predicate marker *i* - that is, in circumstances in which we can be reasonably sure it is used as the abilitative marker rather than one of the other functions detailed above - 511 (or around 87%) of occurrences are preceded by *no* and therefore negated. The same pattern is reflected in taking a look at the negated vs. non-negated uses of the 10 most common verbal collocates of *inap*, as table 6.26 shows. For comparison, the (spoken part of the) BNC gives 3,331 hits for *able to* and only 141 for *unable to*, while a query of *can* + V ends up at 21,572 instances and *can not* + V at only 702 instances. If anything, the pattern here is reversed, suggesting a close link between formal negation on the one side and conceptual inability on the other side.

Verbal collocate	Total	pos	neg
lukim (see)	281	11	271
kisim (get)	156	20	136
kilim (beat, kill)	412	10	122
kaikai (eat)	113	23	90
save (know)	83	5	78
painim (look for, find)	77	4	73
kam (come)	76	3	73
mekim (make)	53	7	46
karim (carry, endure)	51	14	37
wokobaut (walk around)	46	4	42

Table 6.26: Negated to non-negated uses of *inap* with its 10 most common verbal collocates

Therefore, to me, the origin of *inap* as an abilitative construction is likely to be as follows: English *enough* was first borrowed into Tok Pisin as a stand-alone predicate in the construction *Em inap* 'it's enough', in a similar vein as *Em tasol* was adopted as a narrative/discourse marker. From there, it would have

first expanded its function to a modifier in constructions such as examples 151 and 152 above. The next step would have proceeded into two different directions, i.e., two new branching grammaticalization paths for *inap*, while preserving the use as a stand-alone predicate and a modifier. One of these paths leads to the abilitative, while the other leads to the prepositional use when combined with *long*. It is also noteworthy that the abilitative path has been initiated by *inap* alone, but in combination with *no*, i.e. in the negated form.[97] Then, the prepositional path also leads to a functional extension as a conjunction, sometimes still accompanied by *long*. Finally, the abilitative *inap* eventually expands to cover a permissive function. The suggested development is summarized in the schema below:

<div align="center">

inap 'enough'

$\Downarrow$

stand-alone predicate + modifier

$\Downarrow$

abilitative $\Leftarrow$ stand-alone predicate + modifier preposition $\Rightarrow$ preposition

$\Downarrow$

abilitative $\Leftarrow$ SAP + modifier $\Rightarrow$ preposition+conjunction

$\Downarrow$

abilitative+permissive $\Leftarrow$ SAP + modifier $\Rightarrow$ preposition+conjunction

</div>

The concept behind the reanalysis of *no inap* or *inap* as the abilitative lies in the reliance of 'enough' on other syntactic units. English *enough* will serve to illustrate. No matter whether it occurs as a determiner, adverb, or pronoun, it denotes an (in)sufficient quantity of X (to do Y). As a determiner or adverb, the action of Y may be elided, as in *That's enough firewood* or *I can run fast enough*, but an action is still implied. As a pronoun, the insufficient quality may be elided, but the action has to still be present, as in *That's enough to keep us fed for a while*. Of course, an item/quality/etc. is still implied. That this concept might have played a role in Tok Pisin's *inap* as well is reflected by the fact that the construction *no inap long* + V as in the examples below is still very frequent. As Verhaar (1995, 138) notes, "*Inap* as a modal auxiliary is followed by the core constituent, either immediately or with either *long* or *i* (but not both) between it and the core constituent".

---

[97]In this regard, Verhaar (1995, 138) also makes the observation that while *inap* "rarely if ever" is preceded by the predicate marker *i*, "negated *inap* takes *i* wherever required by the rules". Rather than a structural or functional difference, however, this is more likely due to the fact that *i inap* would be phonologically undesirable, as the same vowel would follow itself, which would be hard to perceive.

(167) Mama mi **no inap long** apim taro, em tokim mama.
Mother 1SG NEG ABL PREP? lift taro 3SG tell mother

Mother I cannot lift the taro, he told his mother. (ZC 00710)

(168) Na em wok long slip i stap arere long
CONJ 3SG PROG PROG sleep CONT CONT PREP-L PREP

dispela hap wara, em i **no inap** long muv.
DT place water 3SG PM NEG ABL PREP move

And he was sleeping next to the water, he was unable to move. (ZC 02537)

Once again, we find *long* as part of a construction, similar to complex prepositions PREP + *long*. It is conceivable that like with *wantaim*, *long* here served as a catalyst for grammaticalization.

At the same time, we have two competing constructions coming, once again, from alternative grammaticalization paths. The first of these is *ken*, which, as described above, emerges from the permissive path. This origin is also evident in the fact that the negated form, *no ken*, most commonly functions as the prohibitive rather than a negated abilitative (Wurm and Mühlhäusler 1985, 386). The second competing construction is, of course, *save*, emerging from the KNOW path of grammaticalization.

These three paths show variable degrees of attestation in other languages. Heine and Kuteva (2002, 327) list the concepts of ARRIVE, GET and KNOW as possible other sources of the abilitative. The source of KNOW accounts of course for the abilitative constructions in Tok Pisin, Bislama and Solomon Islands Pijin as based on *save*. The ARRIVE and GET sources seem irrelevant for Tok Pisin. For the first, the concept of arrival is linked to the concept of succeeding at something, which opens up the path to the ability to do something as the capacity of succeeding at it. In fact, SUCCEED is another possible source concept for the abilitative. However, Tok Pisin has no verb equivalent to English *succeed*. The expression *karim kaikai*, literally 'carry fruit' is, akin to English *bear fruit*, sometimes used to indicate that an action was successful or delivered the expected or desired outcome. 'Arrive' in Tok Pisin is *kamap*, 'arrive at' is *kamap long*, neither of which is used in an abilitative function. Instead, it is one of the possible forms of the inchoative (see chapter 6.8.3). Due to its strong link with the concept of beginning and coming into existence, it was likely not a good candidate for grammaticalization as the abilitative through the ARRIVE path.

The path from GET to the abilitative is less easily dismissed for Tok Pisin. *Kisim* 'to get, take, claim, obtain, receive' is a highly frequent verb: at 65,736 occurrences, it is the 25th most common word and third most common verb
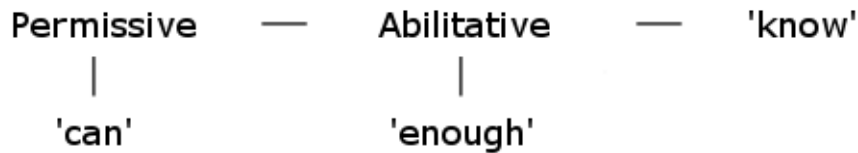
Figure 6.11: Grammaticalization map for the abilitative

in the Z'graggen corpus (after *tok* and *stap*).

While Heine and Kuteva do not list examples of grammaticalization from the permissive to the abilitive - and therefore the path for *ken* - they count the opposite direction to be "a well-researched instance of grammaticalization" (2002, 27), accounting for the path of *inap* from abilitative to permissive. There are, however, many grammaticalization paths which work in both directions, and I see no reason why the path between permissive and abilitative should be an exception to this rule. Once again, we find the competition between several forms for the same function not as a result of a spontaneous demand for a construction or direct super- or substrate influence, but the natural, internal development in Tok Pisin. The different grammaticalization paths are once more summarized in a semantic map below:

Ken, emerging from the 'can' path to cover the permissive and eventually the abilitative, is still the predominant form for the permissive. *Inap*, emerging from the middle part to first the abilitative, is continuing, as per the observations of Siegel and Smith above, to cover the permissive as well. This leaves *save* for analysis.

Any analysis of *save* has to start with its main function, which is that of a habitual marker: as Mühlhäusler (1985, 381) notes, "predicates preceded by *save* denote actions which are performed habitually". Succinct and correct as that summary may be, it does not cover the habitual function of *save* in its entirety, given that the habitual in Tok Pisin is rather more complex than in English. Verhaar (1995, 151) summarizes the habitual functions of *save*:

> [...] English glosses of clauses with the auxiliary *save* are bound to vary a great deal, because English does not have an auxiliary of this kind fitting all contexts. Some English examples are: *Boys **will** be boys*, where *will* expresses what *save* does in Tok Pisin. Another one is *used to*, as in *They **used to** play cricket*, but of this English only has a past tense. In *He **always** shaves in the morning*, what *always* expresses is rendered by *save* in Tok Pisin. Yet another English "parallel" of Tok Pisin *save* is *would*, as in *They would ask*

the strangest questions. Finally, English frequently does not express what is rendered by Tok Pisin *save* at all, as in *We have our meetings at 9 a.m. here,* which in most contexts would express a habit or custom. Tok Pisin would use save to express this.

As complex as it may be functionally, Verhaar also points out that "syntactically, the auxiliary *save* is not very complex", preceding the predicate and being directly followed by the core, without any intervening material - not even the predicate marker *i.* The following two examples show *save* as a habitual marker:

(169) Na   pikinini ia   tok, Em pikinini meri,   wanpela pikinini
      CONJ child   FOC say 3SG child   woman NUM   child
      meri   susa bilong   mi **save** slip long   dispela bet.
      woman sister PREP-P 1SG HAB sleep PREP-L DT   bed
      And the child said - she was a girl - "one of my sisters sleeps in this bed". (ZC 03505)

(170) Em **save** wokim long   banara bilong   em na   malenim tu
      3SG HAB work   PREP bow   PREP-P 3SG CONJ paint   also
      long   pos bilong   haus bilong   em.
      PREP? pole PREP-P house PREP-P 3SG
      He worked on his bow and also painted the poles of his house. (ZC 00815)

Bislama has two ways of denoting the habitual: *save*, which only has a habitual function, and *stap*, which, in addition to the habitual, can also express the progressive Meyerhoff (2013).

(171) Ol SDA oli   no   **save** kakae mit.
      PL SDA AGR NEG HAB eat   meat
      Seventh Day Adventists don't eat meat. (APiCS-Bis)

(172) hem i   **stap** kasem fits
      3SG AGR HAB catch   fits
      She gets (epileptic) fits.

(173) olgeta oli   **stap**   yusum fasin   blong bifo   yet.
      3PL   AGR PROG/HAB use   fashion POSS before yet
      They still do things the old fashioned way. (APiCS-Bis)

For Solomon Islands Pijin, Jourdan (2004, 710) reports that "the verb save acts as a modal indicating habituality and ability", giving the following two examples (annotation added by me):

(174) Hem **save** sevis   long   sande. [her ex. 23]
3SG HAB go-church PREP Sunday

She (usually) goes to church on Sunday.

(175) Pita no **save**   draeva. [her ex. 24]
Peter NEG HAB/ABL drive

Peter cannot drive.

None of the remaining 24 English-lexified contact languages in the APiCS data use *save* in the function of a habitual marker, as table 6.27 shows:

Language	Habitual marker
Early Sranan	*de*
Sranan	*e*
Saramaccan	*ta*
Nengee	*e*
Creolese	*das*
Trinidad Endlish Creole	*doz*
Vincentian Creole	*a*
Jamaican	*de*
Belizean	*de*
San Andreas Creole English	*stody/yuuztu*
Nicaraguan Creole English	*doz/stodi/yuuztu*
Bahamian Creole	*does/is/used to*
Gullah	*da*
African American English	*be*
Krio	*kin*
Ghanaian Pidgin English	*dè*
Nigerian Pidgin	*dè*
Cameroon Pidgin English	*di*
Pichi	*kín (dè)*
Chinese Pidgin English	-
Singlish	*-ing*
Tok Pisin	*save*
Bislama	*save, stap*
Norf'k	*yuus'*
Kriol	*oldei orlas ol-wei ala/yusdu yustu*
Hawai'i Creole	*justu*

Table 6.27: Habitual markers in the 26 English-lexified languages in the APiCS
data

Here, we find one of the instances in which alle three dialects of Melanesian
Pidgin align, yet differ from all remaining English-lexified contact languages.
This suggests that rather than being a later development or innovation, *save*
was already present - at least in form - at an early stage of Melanesian Pidgin.
This proves to be true when examining early data as provided, for instance, by
Mühlhäusler (2003), which shows *save* being used in a list of Pidgin English
expressions as early as 1883[98]:

(176) You **save** this man where he   stop?
       2SG know DT man where 3SG stay
       Do you know where this man stays?

(177) You go house belong   A., you speak him, he   **save** where
       2SG go house PREP-P A. 2SG speak 3SG 3SG know where
       pulumakau he   stap.
       beef   3SG be
       You go to A's house and ask him if he knows where I can get beef.

These two examples already contain a hint as to the commonly assumed ori-
gin of *save* in Portugese *saber* (Lehmann 2015, 37) 'to know': note the two
different spelling *save* and *sabe*, the latter closer to the putative etymon. How-
ever, the form is hardly unique to Tok Pisin and its sister dialects. Rather, it
is, along with items such as *pikanini*, one of the few indisputable elements of
English foreigner talk, South Seas Jargon or Pacific Jargon English (the lines
separating these varieties being fluid), having derived from Atlantic pidgins
and creoles (Keesing 1988, 93; Crowley 1990, 189; Holm 1988, 147). Its devel-
opment was summarized by Aitchison (1989, 165-166) as consisting of three
steps: first, from lexical 'know' to lexical 'know how to' [99]; secondly, from
'know how' to 'be accustomed to' to 'be used to'; finally, from 'be used to' to
a habitual action. In other words, *save* has gone from 'know' to the abilitative
to the habitual.

　　However, it is not fully clear whether *save* was a habitual marker or an
abilitative marker first, or whether there was a meaningful difference in time
when it came to acquiring these two functions. The habitual is already recorded

---

[98]Glossing mine, translation as provided by Mühlhäusler.

[99]Note how this also contains a scope expansion from NPs to VPs, similar to *tasol* in chapter
6.6.2.

as one of the four meanings of *save* in Mihalic's dictionary (1957, 125): "to do often, frequently, habitually; to be wont to, to be in the habit of" in examples such as "*Mi no save giaman.*" 'I do not lie'. However, so is 'to know how to' [100]. Hall's dictionary section also records several meanings or functions of *save*, among them nominal 'knowledge', a transitive verb 'know; have sexual intercourse', as a construction of *save* + verbal complement 'know how to', 'be able to do' and 'do frequently, habitually'. Both entries therefore include both the habitual and abilitative meaning. Furthermore, *save* already had acquired the functions of habitual and/or abilitative marker in early Melanesian varieties such as Early Beach-la-Mar (Holm 1988, 147). As Crowley (1990, 189) notes, however, "the only attested example of pre-verbal *save* from Early Beach-la-Mar is ambiguous between marking ability and habitual aspect".

I suggest that rather than the abilitative or the habitual function preceding one another, *save* showed this ambiguity - absent a clarifying context - from its start as a TMA marker, and it was this ambiguity, along with possible substratum influence, that led it to evolve in two different ways in Tok Pisin and closely related Bislama. While in Tok Pisin, *save* serves as habitual and abilitative marker, it is a habitual, permissive and abilitative marker in Bislama. In the habitual function, it was partly replaced by *stap*, which, like in Tok Pisin, also serves as a continuous marker in Bislama - albeit exclusively pre-verbally.

It is this distinction between *stap* + V in Bislama and V + *i stap* in Tok Pisin that leads Crowley (2004, 217f.) to point out that "in adopting *stap* to express [the] category [of continuous], Tok Pisin had the option of placing the form either before the verb or after the verb" and that "the fact that *i stap* sequences probably already occurred in the language as serial constructions with verbs of direction may have predisposed speakers towards the latter solution rather than the former", explaining the emergence of V + *i stap* for the continuous. This also meant that *save* and *stap* did - for the most part[101] - not occur in the same syntactic slot, with *save* occurring preverbally and *i stap* occurring postverbally.

Substratum influence is unlikely to have played a role for *save* in Tok Pisin. As Mosel (1980, 126) notes, "[t]here are no auxiliaries in Tolai", with "[t]he notions of wish, competence, ability and obligation [being] expressed by full verbs plus noun or verbal noun, plus prepositional phrase or clause". There are various Oceanic languages with preverbal markers of ability or pre-

---

[100]The other two entries under *save* are nominal 'knowledge' and verbal 'to know'.

[101]Verhaar (1995, 114 points out that while "Tok Pisin grammars record *stap* + core verb [...] as a progressive of the core verb [...]", it "seems to be rare" in colloquial Tok Pisin and "virtually nonexistent in texts available in print".

verbal auxiliaries denoting the same, such as Arop-Lokep's *ke*, Kaulong's *ako* and *akonen*, Banoni's *tsi*, Kokota's *boka*, or Vinmavis' *rogulel* (Crowley et al. 2013, 261, 400, 401, 449, 519). However, none of these languages is a major substrate of Tok Pisin, and substrate influence would not have been necessary in establishing its syntactic position, given that both the etymon of *ken* and *inap* would have occurred pre-verbally. The same is true for *save*.

This was not the case in Bislama. As Crowley (2004, 217f.) notes, the doubling of an existential verb before the main verb was "fairly widespread" in Vanuatu vernaculars as a means of signaling both a continuous and a habitual aspect. He takes this as strong indication for substrate influence, which led speakers of Bislama to place *stap* preverbally.   However, the difference in position was not the only significant one. While in Tok Pisin, *i stap* came to gradually expand its function from existential to continuous (and eventually to the progressive), in Bislama, *stap* quickly - or maybe even outright - acquired both the habitual and the continuous aspect. Whether it was used just as a continuous marker first and expanded to territory previously held by habitual *save* or whether it was used first as an alternative to habitual *save* remains unclear. In any case, the ambiguity between abilitative and habitual *save* was resolved in Bislama by introducing habitual *stap*.

Tok Pisin eventually resolved the ambiguity differently. Instead of expanding *i stap* to the habitual, it introduced alternative abilitative and permissive markers: *ken* and *inap*. As of their introduction, *save* as an abilitative or permissive marker has become rare. Mühlhäusler points out that "[b]ecause of its ambiguity this construction is little used" in preference for analytic constructions (1985, 387). This seems to hold true in the available data from the Z'graggen corpus. In a sample of 600 occurrences of *save*, 350 were in a lexical function ('to know'), 206 were in a habitual function, and even in a generous reading, only 44 could possibly be ascribed to an abilitative function, and none could be ascribed a permissive function. In addition, many of the latter could also be parsed as the habitual rather than the abilitative, depending on the context. Such low occurrence numbers, especially in light of the high frequency of *save* in other functions, indicate a strongly diminished abilitative function of *save*.

We can therefore summarize *save*'s grammaticalization path as follows (in Tok Pisin):

$$saber \Rightarrow \text{Portugese lexical item}$$
$$\Downarrow$$
$$save \Rightarrow \text{Pacific Foreigner English 'to know'}$$

```
Permissive  ──  Abilitative  ──  Habitual
    |               |       ＼        |
  'can'          'enough'          'know'
```
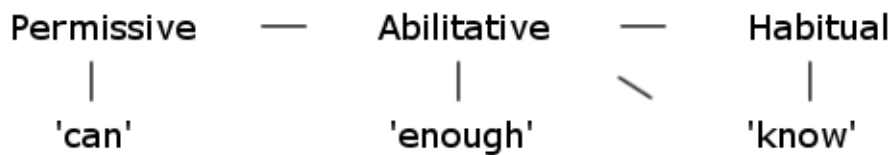
Figure 6.12: Grammaticalization map for the abilitative v2

$$\Downarrow$$

save $\Rightarrow$ 'to know', 'know how to/be able to' (abilitative), 'do frequently'
(habitual)

$$\Downarrow$$

save $\Rightarrow$ 'to know', ('know how to/be able to' (abilitative)), 'do frequently'
(habitual)
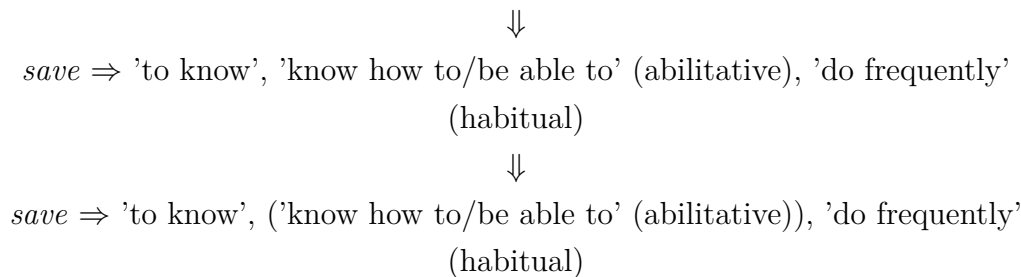
To continue the discussion of the grammaticalization sources of the various abilitative markers above, we finally have the third item, *save*, emerging from the 'know' path. As such, it would hardly have been unusual: Heine and Kuteva (2002, 331) list seven possible source concepts for the habitual, among them CONTINUOUS[102], GO, ITERATIVE, KNOW, LIVE, REMAIN, SIT and USE. Tok Pisin has made use of the KNOW path via *save*. Although it sheds much of its function as an abilitative, in those situations in which it still occurs thus, it is still closely related to "competence in the sense of knowing how to do something" (Wurm and Mühlhäusler 1985, 387), in other words, it mirrors the distinction between skill and physical ability we have also observed in *inaf* and *fitim* in Solomon Islands Pijin above. As Mühlhäusler notes (1985, 387), the habitual and abilitative functions of *save* "are closely related in that one gains one' s competence to perform an action from having performed it regularly or habitually". As such, the path of *save* to the abilitative has led from KNOW to both the abilitative and the habitual:

In Bislama, the path would have begun similarly, but branched out differently as described above.

To summarize, the origin of the various forms competing for the abilitative and permissive functions in Tok Pisin is one more instance of three independent grammaticalization paths converging on one function. One interesting aspect is that while two of these paths are very and fairly common among contact languages, respectively, the other is not. Even so, the seemingly unusual form

---

[102]The continuous being the path for *stap* in Bislama, though the direction is unclear.

still follows well-attested grammaticalization paths, just in a more unusual direction. Secondly, *save* is interesting due to its parallels and differences in closely related Bislama. In the data available, the distribution of the forms still offers some traces of the paths of their individual origins, as we have seen for other constructions before.

6.8.5 Attemptative: *traim*

Whether or not *traim* or *traiim* can be considered an aspectual marker is debatable. It is included here mainly for completive purposes rather than an interesting case study. Mühlhäusler (1985, 378) claims it to be one of the most frequent aspectual markers, and Dutton (1981, 200f.) adds:

> *Traim* (often spelled *traiim*) is used in a way which allows its syntactic interpretation as either a full verb appearing in verbal concatenation or as a verbal marker. This double status can be seen from its position in utterances. It is found either directly preceding the main verb or at the end of a predicate, sometimes even directly after the main verb.

The following two examples show it in use in the Z'graggen corpus. Note how in the first, there is only a single predicate marker for both *traim* and the following verb, while in the second example, both *traim* and *go* are preceded by a predicate marker.

(178) Em i   **traim** kilim meri   nogut long   plenti kain samting.
      3SG PM ATT   kill   woman bad   PREP-C plenty kind thing
      He tried to kill the bad woman with many things. (ZC 01012)

(179) Ol i   **traim** i   go antap, i   go antap nogat tanim
      3PL PM ATT   PM go on-top PM go on-top NEG turn
      kamdaun   gen.
      come-down again
      They try to go on top, cannot go on top, turn and come down again.
      (ZC 00487)

In Bislama, we find the equivalent *traem* can occur in three different variations, either with *bilong* to introduce the following verb, with *long* to introduce the following verb, or without either of the two[103] (Crowley 2004, 180):

---

[103]Glossing mine, translation by Crowley.

(180) Hem i **traem** wokem.
3SG PM try   do
She tried to do it.

(181) Hem i **traem** bilong/long wokem.
3SG PM try   SUB   do
She tried to do it.

In Solomon Islands Pijin[104]:

(182) Had lelebet ... bikos   mi spikim Pijin fastaem, so mi **traem**
Hard little-bit ... because 1SG speak Pijin first   so 1SG try
go [...]
go [...]
It was a little hard ... because I spoke Pijin first, so I tried to speak
[...]. (Emerine 2009, 128f.)

(183) Hem sei: mifala   askem magic man fo mekem wanfala magic
3SG say 1PL-EXCL ask   magic man to make   NUM   magic
wea save kasim Jack sapos   Jack **traem** kilim faestaem
REL know get   Jack COND Jack try   kill   first
He said: we asked a magic man to craft a spell that would get Jack if
Jack tried to kill us first. (Maggio 2018, 70)

In any case, *traim* is most certainly directly derived from English, with no
peculiar grammaticalization path employed or heavy substrate influence to be
gleaned.

6.9 Passive

So far, I have dealt mostly with instances of innovation and grammaticalization
in which a grammatical concept was expressed through either the introduction
of a new formal construction or through reanalysis of an existing construction.
However, not all situations of functional demand were solved by the speakers of
Tok Pisin in this way. The passive, or rather, the lack of a passive construction,
is one such example. According to Verhaar (1995, 334), "Tok Pisin has no
verbal voice; the language has no passive", but instead possesses other means
of focusing nonagent participants. One of these ways is to front the object, as
can be seen in the examples below:

---

[104]Translation provided by source, glossing mine for first; both mine for second.

(184) **Dispela pasin** mipela   i   noken lusim,   mipela   i
DT   custom 1PL-INCL PM can-not abandon, 1PL-INCL PM
mas wokim bihainim yet.
must do   follow   still

This custom we cannot abandon, we must still follow it. (ZC 01567)

(185) Na   lapun meri   tok, Ah **dispela ples** yu lukim, i   nogat
CONJ old   woman say, Ah DT   place 2SG see   PM not-have
man i   stap, em kokonas tasol.
man PM PROG 3SG coconut only

And the old woman said, ah this place you see, it has no people, it has
only coconuts. (ZC 03611)

(186) **Em** yupela kolim Yesus na   mipela   kolim Misil.
3SG 2PL   call   Jesus CONJ 1PL-INCL call   Misil

You call him Jesus and we call him Misil. (ZC 00084)

(187) **Em** mipela   i   no   harim.
3SG 1PL-INCL PM NEG hear

We have not heard it. (ZC 00624)

Verhaar (1995, 334) lists three other possibilities of bringing a nonagent into
focus:

> First, there are reflexive verbs in -*im*. These are transitive morphologically, but
> not semantically. [...] Second, some verb forms have an impersonal subject-
> agent (usually *ol* 'they'), which does not refer to anyone in particular, some-
> thing exceptional for agent subjects. Or the verb form may have a referential
> subject, but one that is low in topicality.

He points out that while reflexive verbs in English can have an object, this
object can only be a reflexive pronoun matching the subject, as in *He killed
himself.* It is neither possible to insert a non-matching pronoun, nor is it
possible to put such a construction into the passive voice without the sen-
tence becoming ungrammatical. He lists *painim*, which basically means 'to
look for/seek' or 'to find', but is often used in less literal constructions such
as *painim bagarap* (literally 'find damage', 'to get damaged') as one of few
examples of reflexive verbs with -*im*, the others being *pilim* ('to feel'), *tanim*
('to turn') and *tantanim* ('to roll'). Example 188 illustrates this fact:

(188) Na   man bilong   em i   **tanim** na   lukim meri   bilong
CONJ man PREP-P 3SG PM turn   CONJ see   woman PREP-P
em i   slip   strong tumas.
3SG PM sleep strong too-much

And her husband turned and saw his wife sleep too deeply. (ZC 03311)

As for impersonal subject-agents, Verhaar (1995, 336) writes that nonreferential use of *ol* "makes it possible to make the subject [...] prominent as new information in context - as a new topic, as the principal focus". This structure is similar to English *they V* as in *They say he's a good guy*, insofar as *they* refers to no particular referents. An example from the Z'graggen corpus:

(189)   Long   tokples   Moro **ol i kolim** Savur.
        PREP language Moro 3PL   PM   call Savur

        In the Moro language they call/which is called Savur. (ZC 04338)

Bislama shows similar structures to Tok Pisin in this regard, as does Pijin:

> There is no passive construction at all in the grammar of Bislama. However, the function of this so-called 'agentless passive' corresponds very closely to the function of the subjectless predicate construction marked with the predicate marker *oli*, i.e.
> *Oli stepem tep blong mi.* [...]
> (Crowley 2004, 121)

> Pijin does not have passive verbs, nor do most of the Solomon Islands vernaculars. This is not usually a major problem in translation when both the goal of the action and the agent or doer of the action are specified, as in 'That man [subject as goal] was arrested by the police [doer].' The purpose of the passive in this sentence is to focus on *that man*. This sentence can be translated into Pijin as: *Disfala man ia, olketa pulis kasholem nao* (This man, the police arrested). (Mugler 1996, 204)

Heine and Kuteva (2002, 23) define the passive marker as "a marker indicating that the action is viewed from the perspective of the recipient or patient of the verb, while the agent is suppressed or demoted". As we have seen above, this can be accomplished without overt passive marking by changing the word order, thereby focusing on the recipient or patient. However, the fact that this strategy exists in Tok Pisin (as well as Bislama and Solomon Islands Pijin) suggests that there was a categorial demand for focusing the recipient/patient. So what other possibilities would there have been to accomplish this function?

Heine and Kuteva (2002, 333) list no less than nine possible source paths for the passive: ANTICAUSATIVE, COMITATIVE, EAT, FALL, GET, PERSPRON THIRD PLURAL, REFLEXIVE and SEE. The first we can rule out, given that to my knowledge, Tok Pisin has no anticausative marker or construction. The COMITATIVE path - which Heine and Kuteva mention might be an areal African phenomenon - seems to be limited to impersonal passives.

This would require either *long* or *wantaim* to occur without a following object in order for syntactic reanalysis to happen. This does not occur with the former, and the latter mostly functions as an adverb meaning 'together' when used without an object, which seems unlikely to provide a syntactic environment conducive to reanalysis as a passive, given that the subjects engaging in the action are usually expressly named either though nouns or pronouns.

For EAT and FALL, they cite similar caution, noting that "the conceptual base of this grammaticalization is not entirely clear", although both "appear to be an instance of a more general process whereby constructions involving certain process verbs are grammaticalized to passive constructions" (2002, 122, 133). Incidentally, this is also true for GET and SEE. In Tok Pisin, the corresponding forms would be *kaikai* 'to eat', *pundaun* 'to fall', *kisim* 'to get' and lukim 'to see'. Some of these are highly frequent words, with *kaikai* at 52,709 occurrences[105], *lukim* at 39,386 and *kisim* at 65,736. However, they are infrequently used in a context conducive to a process reading. *Lukim* and *kaikai*, are, for instance, frequently followed by nominal objects. In order for reanalysis as passive marker to happen, a suitable syntactic environment would have them followed or preceded by another verb, which happens much more infrequently. The most common verb to follow *i kaikai*, for instance, is *mekim* at a total of only 24 occurrences. *Pundaun* rarely combines with other verbs either, with only *kamdaun* being an exception at 257 of a total of 5,828 occurrences. Even if these process verbs are used in close proximity with other verbs, it is usually clear from context that they describe a sequence of events, and the subjects and objects of the action are overtly marked, further reducing the likelihood of a passive or patient-focused reading.

Kisim is the only of these process verbs which lends itself to an occasional passive reading, as in the examples below: The personal pronoun for third person plural already doubles as a plural marker, and might therefore be unlikely to develop into yet another function. Furthermore, it occurs either preceding a noun as the plural marker, or before the predicate marker as a pronoun. As such, it is not in a syntactic position conducive to reanalysis as a passive marker. For the reflexive, Heine and Kuteva (2002, 253) posit an obligatory anticausative stage, which, as mentioned above, does not exist in Tok Pisin.

In summary, Tok Pisin's strategy of fronting the patient/object of a clause to focus the same is taken from neither English nor German, which both have formal means of marking the passive. Instead, speakers of Tok Pisin have opted for purely syntactic means.

---

[105]Albeit including usage as a noun.

6.10 Relativizers: *ia, we* and *husait*

Relativizers in Tok Pisin function akin to relativizers in other languages: as a "morpheme or particle that sets off the restricting clause as a relative clause" (Payne 1997, 326). Sankoff and Brown (1976, 632) describe the prototypical relative-clause construction in Tok Pisin as one in which "the embedded relative is bracketed off from the matrix sentence by the particle *ia*", which would quality *ia* as a relativizer. The initial, or left-hand *ia*, in their examples, occurs "immediately after the head noun", while the second *ia* occurs after the embedded clause it signals. They also note that the modified noun can occur in either subject, complement and circumstantial or oblique position, giving an example for each. This claim is further corroborated by the following examples 190, 191, 192 and 193 from the Z'graggen corpus:

(190) Na   man ia, Mambra i   go antap long   haus tambran
       CONJ man FOC Mambra PM go on-top PREP house ancestor
       bilong   ol   na   lukim pinis   na   tok, Em man **ia**   em
       PREP-P 3PL CONJ see   COMPL CONJ say 3SG man REL 3SG
       i   ronowe   **ia**, i   go i   stap wantaim mipela   na   em
       PM run-away REL PM go PM be   PREP-C 1PL-EXCL CONJ 3SG
       i   kam.
       PM come
       And then this man, Mambra, went on top of the house of the grand-parents and, having looked around, said the man who ran away went to be with us and he came. (ZC 03841)

(191) Q: Wanem pikinini? A: Em pikinini **ia**   em i   go daun **ia**.
       Q: What   child?   A: 3SG child   REL 3SG PM go down REL
       Q: Which child? A: The child who went down. (ZC 00672)

(192) Oh brata   bilong   em **ia**   em i   dai longtaim **ia**, bodi
       Oh brother PREP-P 3SG REL 3SG PM die long-ago REL body
       bilong   em, em nau ol   krai i   stap na   ol   planim.
       PREP-P 3SG 3SG now 3PL cry PM PROG CONJ 3PL bury
       Oh his brother who died long ago, his body, now they were crying and they buried him. (ZC 02932)

(193) Yes em   tanim trausel, meri   go daun   long   solwara **ia**
       Yes 3SG turn   tortoise woman go PREP-D PREP saltwater REL
       em i   tanim trausel **ia**   mipela   kaikai.
       3SG PM turn   tortoise REL 1PL-INCL eat
       Yes it turned the tortoise, the woman went down to the sea which turned the tortoise we eat. (ZC 03782)

As can be seen from the examples above, *ia* bracketing functions not unlike English relativizers such as *that*, except for the fact that *ia* occurs both before and after the embedded clause. Sankoff and Brown (1976, 633) also note that "[a]nother aspect of flexibility in ia-bracketed relatives is the way that the coreferential NP is represented within the embedded relative clause [and that it] appears never to be simply copied as a noun, though sometimes it is represented by a pronoun". This is also the case in the sentences above, with *man* in 190 represented by pronominal *em*, as is the case with *pikinini*, *brata* and *solwara* in 191, 192 and 193, respectively. Another possibility is deletion of the coreferential NP, so that "the head NP has no surface representation in the relative clause" (1976, 633). Example 194 below illustrates such a deletion:

(194) Guguk i   gat dispela kain spia **ia** i   gat mak long bes
       Guguk PM have DT   kind spear REL PM have mark at   base
       ia, i   gat bukbuk ia.
       REL PM have crocodile FOC
       Guguk had this kind of spear which has a mark at the base, it had a
       crocodile. (ZC 03867)

As for the origin of *ia* as a relative marker, Sankoff and Brown (1976, 639) note the following:

> [I]t is as a demonstrative or deictic marker that ia abounds in our data. It seems only a short step to extend the function of a lexical item that has served as an adverb of place to a demonstrative or generalized deictic function. At least, this is a phenomenon common to many languages; cf. Eng. *this here man*, which retains a non-standard connotation, or Fr. *celui-ci* and *celui-là* ('this one' and 'that one' in standard French, from the adverbs ici and la respectively). We should point out that, although the argument in this paragraph assumes a 'place adverbial' origin, with an extension to broader demonstrative or deictic functions (an argument which appears to have historical support in this case [...]), the fact that the two functions are expressed by the same form on the synchronic level, in Tok Pisin as in many other languages, is understandable in terms of the close semantic analogy between the two uses, without assuming any directionality.

Ia, in the spelling *hia*, abounds as a spatial adverbial in the Z'gragggen corpus as well, as does the demonstrative function, often after *dispela*, as examples 195, 196, 198 and 199 show:

254

(195) Yes i   stap **hia**   long   Kaiten.
      Yes PM be   ADV-L PREP-L Kaiten

      Yes I was here at Kaiten. (ZC 00079)

(196) I   go insait **hia**   long   wanem ol i   pret long paia
      PM go PREP-D ADV-L because because 3PL PM afraid PREP fire
      nau.
      ADV-T

      They went inside here because they were afraid of the fire now. (ZC
      04312)

(197) Na   mama tok, Em **hia**.
      CONJ mother say 3SG DEM

      And the mother said, this one here. (ZC 00103)

(198) Q: Husait   i   go was long   garden? A: Em man bilong
      Q: PRN-INT PM go wash PREP-L garden? A: 3SG man PREP-P
      dispela meri   **ia**.
      DT   woman DEM

      Q: Who went to the garden to wash? A: The man of this woman. (ZC
      03728)

(199) Na **dispela** mama **ia**   em i   storim dispela pik na tok,
      Q: CONJ   DT   mother DEM 3SG PM   store   DT pig CONJ
      Yutupela i   go [...]
      say   2PL PM go

      And this mother she stored this pig and said, You go [...] (ZC 01924)

We can therefore assume that the first step in *ia*'s development was from a
spatial adverbial to a demonstrative. In Tok Pisin, however, there is not only
a close semantic analogy between the use of *ia* as a demonstrative and an
adverb of place, but also a close semantic and syntactic analogy between *ia* as
a demonstrative and as a relative marker. Consider examples 200 through 203
below, which can, without context, be parsed both ways:

(200) Yes pikinini **ia**   em i   givim long   lapun man
      Yes child   REL/DEM 3SG PM give PREP old   man
      ia.
      REL/DEM

      Yes, the child who offered to the old man. / Yes, this child offers to
      this old man. (ZC 03057)

(201) Yes, asalai man **ia**   em i   lukim bikpela brata
      Yes, spirit man REL/DEM 3SG PM see   big   brother
      ia.
      REL/DEM

Yes, this spirit saw the big brother. / Yes, the spirit who saw the big brother. (ZC 03474)

(202) Na   pisin man **ia**   em i   no   laik bihainim
CONJ bird man REL/DEM PM NEG want follow woman
meri   **ia** i go   long ples, em tokim meri ia,
REL/DEM PM go PREP-D place 3SG tell woman DEM 1SG
Mi   noken go.
cannot go

And this lyrebird did not want to follow this female to the place, he told this woman, I cannot go./And the lyrebird who did not want to follow this woman to the place, told this female, I cannot go. (ZC 03549)

(203) Na   em dispela man **ia**   em mekim singsing long   sel
CONJ 3SG DT   man REL/DEM 3SG make song   PREP shell
ia,   sel ia i bekim em, Hoi.
REL/DEM shell DEM PM answer 3SG Hello

And this man sung for this shell, this shell answered him "Hello". / And to the man who sung for the shell, the shell answered, "hello". (ZC 03108)

Note that there are, of course, examples in which *ia* is used unambiguously as a relative marker, most obviously when the second *ia* comes after a verb, and can therefore not be a demonstrative (example 204 below). In other occurrences, a parsing as both are syntactically, but not pragmatically possible (example 205):

(204) Orait   tulait nau man **ia**   i   go lukim kandre bilong   em
Alright dark now man DEM PM go see   uncle   PREP-P 3SG
sikau,   na   man ia   em save ia.
wallaby CONJ man REL 3SG know REL

Alright, it was dark now, this man went to see his uncle wallaby and a man whom he knew. (ZC 03685)

(205) Ol   man i   katim garden na   i   olsem   dispela man **ia**
PLM man PM cut   garden CONJ PM like-this DEM   man REL
em i   no   save ol   man **ia** i   go wok garden.
3SG PM NEG know PLM man REL PM go work garden

The men cut the garden and like that, this man who did not know them went to work in the garden. (ZC 02986)

These syntactic and semantic relations make it plausible to trace a development from a spatial adverbial *ia*, to a postposed demonstrative *ia* to the bracketing

structure of *ia* as a relativizer, as Sankoff and Brown (Sankoff and Brown 1976, 663) do:

> Our reconstruction follows simply from this historical account. That is, we propose three stages: (1) the original 'place adverb' ia; (2) extension for use as a postposed deictic or demonstrative; and (3) further extension for general 'bracketing' use, including topic-comment structures, relativization, and cleft sentences. That these uses are semantically and functionally related has been shown [...].

They further relate this development to the creolization process in Tok Pisin. Noting that attestations of *ia*-marked relatives date back further than any significant numbers of creole speakers, they assume that rather than an innovation among fluent second-language speakers of Tok Pisin without further motivation, the expansion from demonstrative to relative marker was due the semantic and functional compatibility on the one hand, and substrate influence from Austronesian languages on the other. With the origin of *ia* satisfyingly explored for the moment, there are at least four other possibilities to express relativization in Tok Pisin, the first being intonation, which will be disregarded here on the account of the audio recordings not being available for analysis. As to points two to four, as Sankoff and Brown (1976, 636) note, there is the possibility of using *wh*-forms for relativization:

> Another possible type of marker is WH-forms-a set of obvious candidates for relativizers in any language, given the relation between relatives and indirect questions as discussed, e.g., in an important paper by Keenan [and] Hull 1973. In Tok Pisin, these are *we* 'where', *husat* 'who', and *wonem* 'what'. In all the complex sentences we have examined, only five use WH.

Two of these options appear in the data of the Z'graggen corpus, namely *we* and *husait*, as examples 206 through 211 below show. There are no instances in the Z'graggen corpus of *wanem* or *wonem* 'what' being used as a relativizer.

(206) Olgeta meri **husait** i karim, em i kilim na kaikai.
All woman REL PM carry 3SG kill CONJ eat

He killed and ate all the women who carried. (ZC 00093)

(207) Q: Husait i givim nem? A: Man **husait** i bin kamapim
Q: PRN-INT PM give name A: Man REL PM PSM find
ol.
3PL

Q: Who gave the name? A: The man who found them. (ZC 00293)

(208) Ol   save biruaim wantok, wantok **husait** em i   mekim wanem
3PL HAB damage family   family   REL   3SG PM make   what
kain kain pasin   long ai bilong   ol.
kind kind custom PREP eye PREP-P 3PL

They used to damage the family, the family which made what kind of custom in front of them. (ZC 04502)

(209) Na   dispela snek em i   ken tanim skin bilong   em olsem
CONJ DT   snake 3SG PM can turn   skin PREP-P 3SG like
man, na   i   ken tanim skin olsem wanem kain samting,
man CONJ PM can turn   skin like   what   kind something
dispela kain **we** i   stap long   dispela graun yet.
DT   kind REL PM be   PREP-L DT   ground still

And this snake can turn skin of it like man, and can turn skin like what thing, that kind which is on ground already. (ZC 00656)

(210) Tupela kisim dispela man **we** i   kam stap wantaim tupela long
Both   get   DT   man REL PM come be   with   both   at
haus?
haus

Both of them got this man who came be with both at the house? (ZC 03783)

(211) (...) man ia   i   go antap na kisim dispela tupela spia **we**
(...) man FOC PM go on-top and get   DT   NUM spear REL
em i   redim i   stap, wanpela bilong susa   bilong
3SG prepare PM   PROG NUM PREP-P sister PREP-P 3SG
em   na   wanpela bilong   tambu   bilong em.
CONJ NUM PREP-P brother-in-law PREP-P 3SG

(...) this man went on top and got these two spears which he was preparing, one from his sister and one from his uncle. (ZC 01330)

In Bislama, subject and object relative clauses are most frequently marked by a relative particle, combined with either a resumptive pronoun or a gap. The APiCS data Meyerhoff (2013) gives the following examples, among others:

(212) Eni wan **we**   i   pas long kot   bae oli   putum hem long
Any one COMP AGR pass in   court IRR AGR put   3SG in
kalabus.
jail

Anyone who appears in court will be put into jail.

(213) Yu, **we**   yu sapotem gavman,   yu   save kasem wok.
2SG COMP 2SG support government 2SG can get   work.

You who support the government, you'll get work.

```
'who'                                       'here'
  |                                           |
  |                                           |
wh-question  —  relative  —  demonstrative
  |
  |
'where'
```
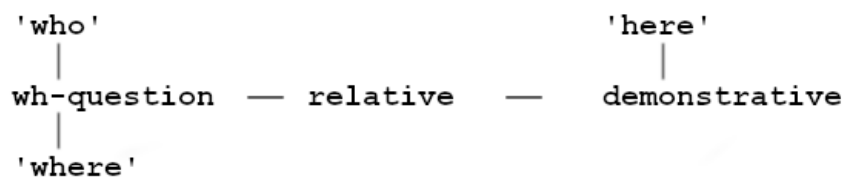
Figure 6.13: Grammaticalization paths for the relative

According to Romaine (1992a, 301), "*wea* is the most common relativizer [in Solomon Islands Pijin], but is associated with certain age group and certain varieties of Pijin". An example would be (Australian Government, 2011):

(214)   Kustoms: Disfala dipatment   hemi lukaftarem olketa samting wea
        Customs: DT   department 3SG look-after all   things REL
        pipol wea travol olketa tekem kam o   tekem aot long
        people REL travel 3PL   take   come CONJ take   out PREP
        kandere.
        country
        Customs: This department looks after all things which people who
        travel take in or out of the country.

Another form "parallel to *husat* [exists] in Solomin Islands Pijin, namely *hu*", which Huebner and Horoi (1979, 183 note "may be used only with human nouns", while "*wea* may be used with either persons or things". They also note variation between retaining and omitting pronouns in both subject and object relative clauses. As such, their example translations for the English sentence *I saw the man you hit* include four possibilities, *Mi lukim man yu hitim hem*, *Mi lukim man yu hitim*, *Mi lukim man hu yu hitim hem* and *Mi lukim man hu yu hitim*.

We therefore find a good deal of variation in the marking of relative clauses in all three Melanesian dialects. However, in Tok Pisin at least, the different variants are not equal. Although *we* and *husait* do occur as relative markers, they are much more rare than *ia* bracketing, and are somewhat limited in their scope. This is, once more, an indication of their origin. Heine and Kuteva (2002, 335) name three different source paths for the relative, among them DEMONSTRATIVE, HERE and W-QUESTION. As shown above, *ia*-bracketing evolved from the HERE path via the DEMONSTRATIVE path to RELATIVE. Both *we* and *husait* evolved from the path of W-QUESTION. The path is in figure 6.13. More specifically, *we* evolved to a relative marker

from the spatial adverb *we* 'where', as used in questions. This is still evident in that it is the preferred option of relative marking for head nouns with spatial semantics, such as *ples* 'place' or *hap* 'place'. *Ia* bracketing only occurs after *ples* and *hap* 16 and 2 times, respectively, in a limited sample[106]. In a similarly limited sample with animate head nouns, however, *ia* bracketing is strongly preferred. *Man* and *meri* showed 113 and 50 occurrences, respectively, of being followed by an *ia*-bracketed relative clause. For contrast, in an unlimited sample, *man* and *meri* are followed by *we* as a relativizer only 11 and 9 times, whereas *ples* and *hap* are followed by a relative clause introduced by *we* 268 and 65 times, respectively. *Husait*, in the entire Z'graggen corpus, introduces a relative clause for *man* and *meri* 24 and 21 times, but never for *ples* or *hap*. Figure 6.14 below illustrates this distribution:
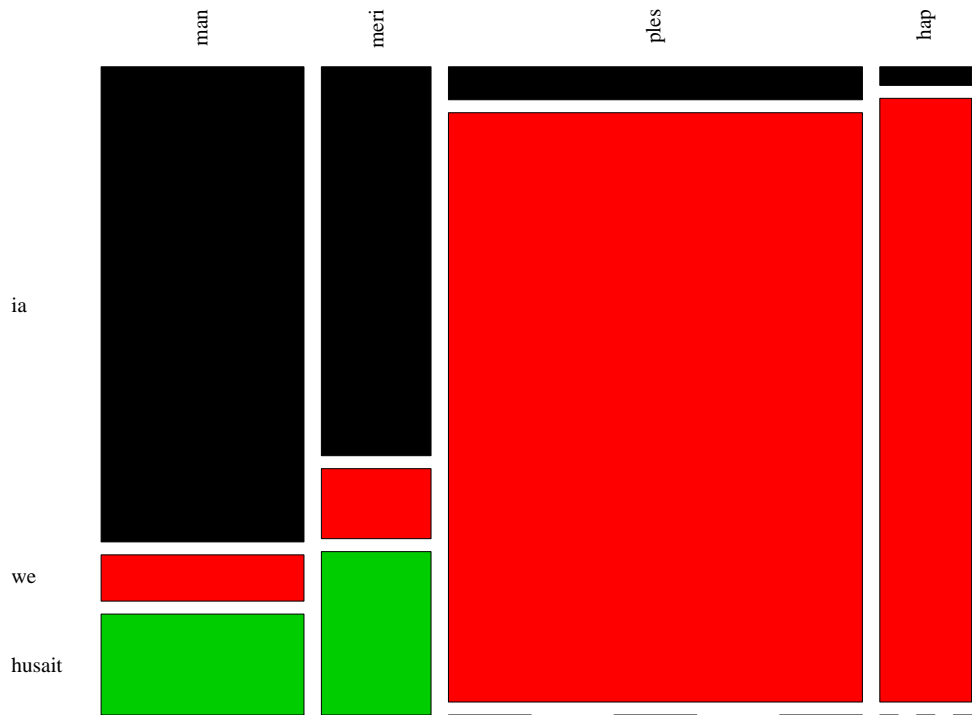
Figure 6.14: Distribution of relativizers across four nouns

Siegel (1985, 528) also remarks that "we is [...] restricted to relative clauses with non-human heads or ones which have an adverbial function". *Husait* also occurs much more frequently as an interrogative pronoun rather than as a relative marker. If it does, however, it can freely occur with animate subjects, likely due to the fact that the answer to its question is usually a subject and

---

[106]Sample consists of all instances in the Z'graggen corpus in which *ples* or *hap* is followed by a relative clause of one two three constituents.

therefore more likely to be animate. On the other hand, it never occurs as a relative marker after *ples*, *hap* or *graun*. In other words, *we* is the equivalent for English *where*, while *husait* is the equivalent of English *who*, which is hardly surprising given that these two English words are their respective etymons, and have gone through the same development from adverbial to relative marker in English. *Ia* bracketing, on the other hand, has no such restriction, and can fulfil the functions of *who*, *what*, *which* and *where*. This is likely one of the reasons for its spread over its alternatives. In summary, we see once again that the origin of the variants for a particular function determines their variation, and that a variant which has no close ties to its original semantic properties has a competitive advantage over those that still exhibit closer ties.

6.11 Complementizers: *(bi)long, olsem, na, we and zero*

Tok Pisin has no less than five possible ways to express complementization, including *(bi)long, olsem, na, we* and zero marking. All five complementization strategies are present in the Z'graggen corpus, as examples 215 through 225 will show. For *long*:

(215) Na   nau i   no inap **long** bai yumi   kisim bek.
       CONJ now PM NEG ABL COMP FUT 1PL-INCL get   back

       And now we will not be able to get it back. (ZC 01289)

(216) Supos Malanga i   tok **long** mekim krismas   long
       COND Malanga PM say COMP make   christmas PREP-L
       nambawan basis, orait   ol   i   ken harim.
       NUM   base   alright 3PL PM can hear

       If Malanga says to celebrate Christmas at the first base, alright they
       can hear. (ZC 04521)

For *olsem*:

(217) Na   man bilong em i   no   save **olsem** em ol famili
       CONJ man PREP-P 3SG PM NEG know COMP 3SG PL family
       bilong   meri bilong   em, em ting em ol pis.
       PREP-P woman PREP-P 3SG 3SG think 3SG PL fish

       And the man did not know that it was the family of his wife, he thought
       it was fishes. (ZC 03601)

(218) Em tok, orait yu tok **olsem** i nogat kaikai.
3SG say alright 2SG say COMP PM not-have food

He said, alright, you say that you have no food. (ZC 00985)

For *na*:

(219) Gutpela **na** yu kam.
Good COMP 2SG come

It is good that you come. (ZC 01130)

(220) Na em tok, Gutpela, sori gutpela **na** yu kam, mi
CONJ 3SG say good EXCL good COMP 2SG come 1SG
stap.
stay

And she said, good, oh dear, it is good that you come, I stay. (ZC 00299)

For *we*:

(221) [...] i kam na em i painim hap ples **we** em i
[...] PM come CONJ 3SG PM find part place COMP 3SG PM
bin slip long en, aninit long lip yam bilong Yukuru
PST sleep PREP-L 3SG PREP-L PREP plant yam of Yukuru
[...]
[...]
[...] he came and he found the place that he slept in, underneath
Yukuru's yam plants [...] (ZC 00668)

(222) Na em tok, Maski, mi traim painim ol pren, ol man
CONJ 3SG say not-worry 1SG try find PL friend PL man
we em i salim.
COMP 3SG PM send

And he said, don't worry, I will try to find friends, men **that** he sent.
(ZC 02523)

For zero marking:

(223) Na Kumnagi em i dai pinis longtaim liklik, na tupela
CONJ Kumnagi 3SG PM die COMPL time small CONJ DT
meri ia i no save Ø em i dai.
woman FOC PM NEG know Ø 3SG PM die

And Kumnagi had died a short while ago, and both women did not
know he had died. (ZC 03032)

(224)   [...] na   em i   save Ø baimbai   bikbrata bilong em i
        [...] CONJ 3SG PM know Ø big-brother PREP-P 3SG   PM want
        laik stil   mekim woa   long em   ia.
        steal make war   PREP 3SG FOC

        [...] and he knew that his brother still wanted to steal from him and
        make war against him. (ZC 00856)

For *bilong*:

(225)   Dispela liklik brata   ia   em i   save long   olgeta
        DT   small brother FOC 3SG PM know PREP everything
        samting, save **bilong** em i   winim gen   bikbrata bilong
        PRN   know COMP 3SG beat again big-brother PREP-P 3SG
        em. (ZC 00903)

        This little brother knew everything, he knew that he had beaten his
        big brother again.

Due to both *bilong* and *long* being used as prepositions, it is possible that
bilong's function as a complementizer shares the origin of *long* as a comple-
mentizer as described above. However, there are also factors suggesting that
the use of *bilong* and *long* as complementizers originated separately. Since
complementization is a relatively complex syntactic phenomenon, it is un-
likely to have arisen when *bilong* was still the only preposition and *long* had
not become established yet. As a complementizer, *bilong* has its origin in con-
structions that feature *save* and *tok* not as a verb, but as a noun, such as *save
bilong ol* 'knowledge of them' as in English *know of (someone)* and *know to
(do something)*. Example 226 below show such a instances of nominal *save*
and *tok*:

(226)   Na   yu i   gat bikpela save   **bilong** wokim olgeta
        CONJ 2SG PM have big   knowledge PREP-P build   everything
        samting.
        PRN
        And you have great knowledge of building all kinds of things. (ZC
        00432)

Smith (2002b, 161) mentions a further complementizer in *sapos*, but notes
that "the use of [...][sapos]" in the role of complementizer seems to be de-
clining. That is certainly true for *sapos*, and does not seem to be a recent
development. Of the 77 occurrences of *sapos* in the Z'graggen corpus, not a

single one occurs in the role of complementizer. It is therefore disregarded in this analysis.

For some languages, variation in the use of complementizers is dependent on lexical features of the verbs they are used with, such as different complementization strategies being used with verbs of knowing and verbs of speaking (for examples including Angolar, Vincentian Creole and Batavia Creole, see chapter 96 of the APICS). However, there is no strong universal tendency among the languages surveyed in the APICs data to use fundamentally different complementizing strategies for verbs of knowing and verbs of speaking, with 15 of them using the same strategies and 10 showing some difference, as table 6.28 shows.[107]

know+	know+	know+	Language	speak+	speak+	speak+
NC	SAY	OTH		NC	SAY	OTH
X	X	X	Early Sranan	X	X	X
X	X	X	Sranan	X	X	X
X	X		Saramaccan	X	X	X
	X		Nengee		X	
X	X	X	Creolese		X	X
	X	X	Trinidad English Creole		X	X
X		X	Vincentian Creole	X	X	X
X	X		Jamaican	X	X	
X		X	Belizean	X		X
	X		San Andreas Creole English	X	X	
X	X	X	Nicaraguan Creole English	X	X	X
X	X	X	Bahamian Creole	X	X	X
X	X		Gullah	X	X	
X			African American English	X	X	
	X		Krio		X	
X	X		Ghanaian Pidgin English	X	X	
X	X		Nigerian Pidgin	X	X	
	X		Cameroon Pidgin English		X	
	X		Pichi	X	X	
		X	Chinese Pidgin English	X		
X		X	Singlish	X		X
X	X	X	Tok Pisin	X		X

---

[107]It remains unclear into which category Norf'k falls, given that the APiCS data does not provide information on its complementizing strategy for verbs of knowing.

X	X	X	Bislama	X	X	
???	???	???	Norf'k	X		
X			Kriol	X		
X			Hawai'i Creole	X		X

Table 6.28: Complementizer with verbs of speaking and knowing in the 26 English-lexified APiCS languages

Tok Pisin's variation in this regard can also not be reduced to a difference between verbs of knowing and speaking, as the examples above show: *olsem*, for instance, is used with both, as is *long*. A look at Tok Pisin's sister dialects does not illuminate the matter either. Like Tok Pisin, Bislama shows all variations of complementizers, as examples 227 to 229 show (Meyerhoff 2013):

(227) mi   save ∅ man ia   i   nogud   nating
      1SG know ∅ man DEF AGR no.good nothing
      I know that man's completely awful.

(228) mi   harem **olsem** mi   sore long pikinini ia
      1SG feel   COMP 1SG sorry for   child   DEF
      I felt like I was sorry for the child.

(229) Oli   save **se**   wanem we   gavman   hemi   mekem
      AGR know COMP what   COMP government 3SG.AGR make
      ino   stret.
      AGR.NEG straight
      They know that what the government is doing is wrong.

Meyerhoff (2013) notes in this regard that "'know' strongly favours *se*", while there are semantic constraints on *olsem*. The complementizer *se* is unique among the Melanesian dialects. The form's origin is somewhat disputed:

> Only Bislama has the complemetizer *se*, derived from the verb of the same form meaning 'say' (Crowley 1989, 1990). Keesing (1988: 50) says this feature was established by the 1870s, but his examples only show *say* being used as a main verb. Crowley (1990: 270) notes that there is no record of *se* being used with a grammatical role from the 1920s or before. Its modern usage has parallels in the substrate languages but also results from contact with the French form *c'est* (Crowley 1989: 206), as well as from universal tendencies to grammaticalize such words as complementizers. Siegel (2008, 194f.)

In Solomon Islands Pijin, complements are marked by *dat*, although Huebner (1979, 182) notes that "there are many areas in the Solomons where *dat* is not used". Speakers in those areas instead use complementizers from local languages or use intonation to convey the same function. For those that use *dat*, it is similar in position and function to English *that*, as the example from Huebner (1979, 182) below shows[108]:

(230)   Mi   taiem hem **dat**   Bili   hem i   stap long   Kukum.
          1SG tell   3SG COMP Billy 3SG PM live   PREP-L Kukum
          I told him that Billy lives in Kukum.

Where, then, do all of these complementizers come from? Heine and Kuteva (2002, 329) list no less than eight possible source paths for complementizers: ALLATIVE, DEMONSTRATIVE, MATTER, W-QUESTION, RELATIVE, RESEMBLE, SAY and THING. Of the five marked complementization strategies, only one falls into any of these categories, namely *we*, which also may be used as a relativizer (see chapter 6.10). The other four - *long, bilong, olsem* and *na* have no such connection. Therefore, they must either have originated via a new - or rather, a previously unobserved - grammaticalization path, or have come up through syntactic reanalysis rather than a semantic connection. For *long, olsem* and *na*, Woolford (1981) has claimed just such a process, claiming that "[a] preposition, an adverb, and a conjunction that occur in positions immediately preceding complement clauses have been reanalyzed as complementizers". The preposition, adverb and conjunction she refers to are *long, olsem* and *na*, the first occurring as a preposition in the types of sentence below[109]

(231)   Mi   amamas **long**   bekim pas   yu   bin   raitim mi   long   bipo.
          1SG happy   PREP answer letter 2SG PST write   1SG PREP before
          I am pleased (for) to answer the letter you wrote to me before.

According to Woolford, speakers of Tok Pisin would have unanimously analyzed such sentences as consisting of a NP, a VP and a PP before World War II, while syntactic reanalysis led some of them to analyze it as NP, VP and COMP later on, leading to the introduction of a complement rule into the language. She (1981, 133) sees the origin of *olsem* in a similar structure, the complementizer arising through syntactic reanalysis of adverbial *olsem* in a suitable position. One of her examples is reproduced below[110]:

---

[108]Translation by Huebner, glossing mine.
[109]Translation by original source, glossing mine.
[110]Translation by Woolford, glossing mine.

(232) Yu   no   ken ting **olsem**   mipela   i   lusim tingting
2SG NEG can think ADV/COMP 1PL-INCL PM lose   thought
long   yu pinis.
PREP-P 2SG COMPL

You must not think/like we have forgotten you.

Similar observations can be made for *na*, which, as Woolford (1981, 135) notes, "appears in a position just preceding clauses" and is thus a viable candidate for syntactic reanalysis as well, as the example below shows:

(233) Na   husait tokim yu **na yu kam**   long dispela garden   bilong
CONJ PRN   tell   2SG CONJ/COMP 2SG come   PREP-P DT
mi?
garden PREP-P 1SG

And who told you to come/and you come to my garden?

She sets down the rules, or tendencies, of variation between *long, olsem, na* and *we* as follows:

1. *Long* is obligatorily present before the complement clauses of certain verbs, but *olsem, na* and *we* are never obligatory.
This is predicted by the fact that prepositions are not deletable, but complementizers are. When *long* is a preposition, it is obligatorily present; but when it is a complementizer, it is optional. The deletability of *long* is determined by the preceding verb because the selection for PP or $\overline{S}$ complements is marked on individual verbs in the lexicon.
2. If *long* is not obligatory, then *long, olsem* and *na* are interchangeable. This is predicted by the fact that if the structure of the complement is $\overline{S}$[COMP S], any one of the complementizers may be chosen to fill the complementizer slot (except the relative clause complementizer *we*). If, on the other hand, the structure of the complement is $_{pp}$[P S], then only the preposition *long* may appear, and as a preposition it is obligatory.
3. There is a great deal of variation among individual speakers as to which verbs take which complement types.
This is predicted if verbs are subcategorized in the lexicon for complement type because things marked in the lexicon are notoriously variable among speakers.

In regards to her first point, note that there is no difference in form between *long* being used as a preposition and it being used as a complementizer. The difference only becomes evident when the lexical disposition of a verb towards either a prepositional phrase or a complement is taken into account. Woolford cautions that her model is based on little data, but it predicts this data very

well. I have tested this model on samples from the Z'graggen corpus in the following three steps:

- Does *long* occur categorically in the complementizer position for some verbs, optionally for others and never for yet others?

- Is there any discernible grammatical or other difference between *long, olsem* and *na* in the other cases?

- Does *we* appear only as a relative clause complementizer?

To answer the first question, I drew four samples of 200 instances of four verbs with complement clauses, namely *laik, tokim, ting* and *save*. I was unable to confirm that *long* is obligatorily present for any of them, given that there was more variety, as figure 6.15 shows. The numbers in the samples indicate that
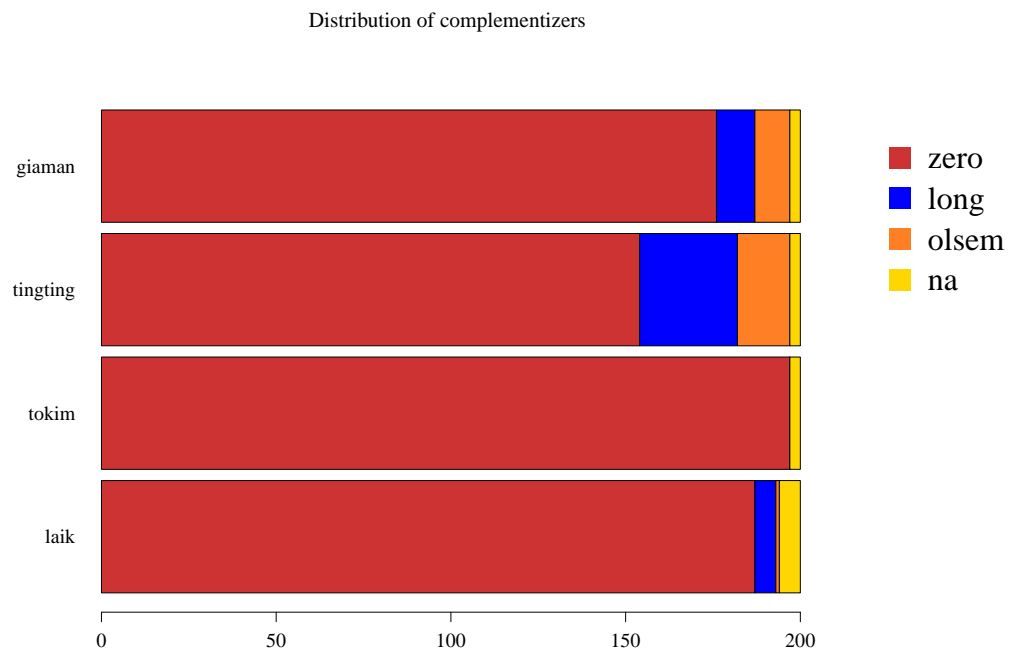
Distribution of complementizers

Figure 6.15: Distribution of complementizers

by far the most frequent option is zero marking for all four verbs, while *na* is the least common overall. However, for three of them, all complementizers do occur. If Woolford's model is correct, that would mean that *long* is not obligatory for any of these verbs, and that the various forms of complementization can occur freely. However, we see that *laik* still takes all three marked types

of complementizer, including *long*, even though in Woolford's sample, *laik* was said to "never have been preceded by *long*, either in the old data from Hall (1943) or in the recent data". Still, the second part of the model holds up on its predictive value. On a possible extension of the sampled verbs, the first part of the model still holds up as well - it has, at least, not been disproven. In addition, there is no discernible grammatical, syntactic or pragmatic difference between the samples in which *olsem*, *na* and *long* occurred respectively, so the variation seems to indeed be free.

As far as *we* goes, there are too few instances of it being used as a complementizer to make a sweeping statement about its status. For what it is worth, however, when it occurs as a complementizer in the Z'graggen corpus, it always occurs within a relative clause. These facts seem to be another indication that Woolford is right in attributing the rise of complementizers to syntactic reanalysis: if there were a semantic component, we would expect there to be a starker contrast between the various uses, as we have seen for other constructions such as the conjunctions or TMA markers. The same is true for complementizing *we* only occurring within relative clauses: it is still tied to the syntactic position it serves as a relativizer, even when used as a complementizer.

Woolford (1981, 137) closes her analysis by observing that "[a]ll of the changes involved in the development of a complementizer system in Tok Pisin are quite ordinary processes of language change" and that "[t]here is nothing involved that is unique to creolization". I can only agree that process-wise, as for most of the other constructions surveyed above, this seems entirely true to me.

6.12 Reflexive: *yet*

It is important for the analysis below to distinguish between reflexive pronouns and intensive pronouns, given that they frequently take the same form within a language, as they do in English and Tok Pisin. Their function, however, is different. While reflexive pronouns are simply pronouns that refer to a syntactically close antecedent (usually a NP within the same clause), intensive pronouns add emphasis to a statement. The latter therefore do not fulfil a syntactically required function for the sentence to become or remain grammatical and are optional. In English, the difference becomes clear in sentences such as *I hurt myself*, in which *myself* is a reflexive pronoun, and *I did it myself*, in which *myself* is a intensive pronoun. In German, for instance, the

distinction is grammaticalized in pronouns such as *sich* and *selbst*, the former being a reflexive pronoun and the latter being an intensive one. König and Gast (2002, 2) put the distinction as follows:

> [Reflexive pronouns] indicate that two participants in the relevant situations [...] have the same referent, i.e. they identify the same person. Such co-reference of self-forms with an antecedent (or binding of a reflexive pronoun by an antecedent), typically the subject of the same clause, is usually represented by assigning the same referential index to antecedent and reflexive (Fredi fancies himself). [Intensive pronouns], by contrast, do not express such co-reference or binding. These forms are invariably phonologically prominent, i.e. they are focused and therefore stressed. The semantic effect of such focusing is the evoking of alternatives. In other words, what [intensive pronouns] do is to bring alternatives to the value denoted by the preceding noun phrase into the discussion.

Mühlhäusler also cautions that "reflexivity in Tok Pisin has not yet been well studied [and] that not all reflexives in English are translated by a construction containing a reflexive pronoun in Tok Pisin", whose speakers often opt for transitive verbs without an overt object to translate relative concepts (Mühlhäusler 2013, 19). The following examples show two cases in which the reflexive pronoun does, indeed, appear overtly in the form of *yet*, postposed after the respective personal pronoun:

(234) Ah maski, bai mi kilim **mi yet**, mi yet mi rong
Ah not-worry FUT 1SG kill 1SG REFL 1SG alone 1SG wrong
na tupela meri bilong mi i lus.
CONJ DT woman PREP-P 1SG PM loose
Ah don't worry, I will kill myself, I am alone I was wrong and I have lost both my women. (ZC 03749)

(235) Em redim kaikai pinis orait em bilaisim **em yet** nau.
3SG prepare food COMP alright 3SG paint 3SG REFL now
He finished preparing the food, alright, and then he painted **himself**. (ZC 03788)

Bislama has multiple ways of expressing the reflexive, including the postposed quantifier *wan*, postposed *nomo* 'only' and postposed *bakegen* 'again' as well as the verbal suffix *-bak*. It does not possess the construction with *yet*, however. As Crowley (2004, 58) notes, these strategies can even all be employed at once, as example 237 below shows:[111]

---

[111]Translation mine, glossing by Crowley.

(236) Mi luk mi **wan** long glas.
1SG look 1SG REFL PREP-L mirror

I looked at myself in the mirror.

(237) Hem i kilim-**bak** hem **wan nomo**.
3SG PM kill-REFL 3SG NUM only

He killed himself.

Solomon Islands Pijin employs the form *seleva* in postposition to the pronoun in order to form reflexive pronouns, resulting in the forms *mi seleva*, *iu seleva*, *hem seleva* and *olketa seleva* for *myself, yourself, himself/herself/themself* and *themselves*, respectively (Regional Assistance Mission to Solomon Islands 2011, 8). Like Bislama, it exhibits no construction with *yet* for this function. The following examples (Huebner and Horoi 1979, 199,212) show reflexivity in use [112]:

(238) Fo taem yu katem **yu seleva**: [...]
PREP time 2SG cut 2SG REFL [...]

In case you cut yourself: [...]

(239) Peim samting fo **mi seleva**.
Find something PREP 1SG REFL

Find something for myself.

Language	Type 1[113]	Type 2	Type 3	Type 4	Type 5	Type 6
Early Sranan	X			X	X	
Sranan					X	
Saramaccan				X	X	
Nengee				X	X	
Creolese				X	X	
Trinidad English Creole					X	
Vincentian Creole	X					
Jamaican					X	
Belizean					X	
San Andreas Creole English					X	
Nicaraguan Creole English				X	X	
Bahamian Creole				X	X	

---

[112]Translation and glossing mine.

[113]Type 1 = Ordinary anaphoric pronoun; Type 2 = Implicit expression; Type 3 = Reflexive marking on the verb; Type 4 = Reflexive pronoun with 'body' or body-part; Type 5 = Compound reflexive pronoun with emphasizer; Type 6 = Dedicated reflexive pronoun

Gullah					X	
African American English					X	
Krio					X	
Ghanaian Pidgin English				X	X	
Nigerian Pidgin				X	X	
Cameroon Pidgin English				X	X	
Pichi				X	X	
Chinese Pidgin English					X	
Singlish				X		
Tok Pisin					X	
Bislama	X					
Norf'k	X				X	
Kriol					X	X
Hawai'i Creole	X				X	

Table 6.29: Reflexive constructions in the 26 English-lexified APiCS languages

Once again, we find that Solomon Islands Pijin reflects the English superstrate most closely among the three Melanesian dialects in terms of structure. As for the rest of the English-lexified creoles covered in APICs data, they employ a similar structure, but not a similar form, as can be seen from table 6.29. Tok Pisin is hardly unusual in using the type "compound reflexive pronouns with emphasizer" as a strategy for forming reflexive pronouns.

Language	Marker Type 4	Marker Type 5	Marker Type 6
Early Sranan	*skin* (<E skin)	*srefi* (< E self)	
Sranan		*ensrefi* (< E self)	
Saramaccan	*sinkií* (< E skin)	*-seéi* (< E self)	
Nengee	*sikin* (< E skin)	*seefi* (< E self)	
Creolese	*skin* (< E skin)	*self* (< E self)	
Trinidad English Creole		*-self* (< E self)	
Vincentian Creole	*brein* (< E brain)	*-self* (< E self)	
Jamaican		*-self* (< E self)	
Belizean		*sɛf* (< E self)	
San Andreas Creole English		*-self* (< E self)	
Nicaraguan Creole English	*hed/skin* (< E head, skin)	*-self* (< E self)	
Bahamian Creole	*head* (< E head)	*-self* (< E self)	

Gullah		*self* (< E self)	
African American English		*-self* (< E self)	
Krio		*ĩsɛf* (< E self)	
Ghanaian Pidgin English	*bɔdi* (< E body)	*-sɛf* (< E self)	
Nigerian Pidgin	*bòdi* (< E body)	*-séf* (< E self)	
Cameroon Pidgin English	*bodi* (< E body)	*sef* (< E self)	
Pichi	*hed, skin* (< E hed, skin)	*sɛf* (< E self)	
Chinese Pidgin English		*-self* (< E self)	
Singlish		*-self* (< E self)	
Tok Pisin		*yet* (< E yet)	
Bislama			
Norf'k		*-self* (< E self)	
Kriol			*(my)jelb*
Hawai'i Creole		*-seof* (< E self)	

Table 6.30: Reflexive markers in the 26 English-lexified APiCS languages

Indeed, it seems the preferred strategy among the languages surveyed. However, once we consider the actual structure employed to form this compound pronoun, it becomes evident once again that Tok Pisin deviates from the common order (see table 6.30). While - invariably - all of the other contact languages adopt a form of English *self* as the reflexive marker or reflexive pronominal suffix, Tok Pisin opts for *yet*. Once more, we find ourselves looking for the source of a deviating form. Heine and Kuteva (2002, 335) list the following three concepts as sources for the reflexive: BODY, HEAD and INTENSIVE-REFL. The first two can be safely discarded here as the source for Tok Pisin's pronoun+ *yet* reflexive strategy, given that *yet* has no etymological or other link to body and head terms. This leaves us with the intensive-reflexive, suggesting that *yet* might have filled that function before.

Brenninkmeyer (1924, 14) reported that "the reflexive pronoun is the same as the personal pronoun, and can only be recognized by its intonation". Mihalic (1957, 159) has 'yet' and *i-no yet* for 'not yet'. While Hall (1943a, 101) records the meanings of *jet*, as he spells it, as adverbial 'yet, still' in the dictionary section of his work, he also notes the following for the grammar section (1943a, 11):

> The reflexive pronouns of European languages are expressed in two ways by Neo-Melanesian speakers, namely:

A. Most commonly by *yet* after a personal pronoun, e. g., *mi yet* "I myself"
yupela yet "you yourselves"
Kamda yet i-paitim em yet. "The carpenter hit himself."
B. Another method in some regions is that of using *bilong* with a personal pronoun ; at times *bilong* and *yet* are used together, e. g., *Bilong em (yet), em i-kam.* "He came of his own free will." *Mi no katim yu, bilong yu katim yu.* "I did not cut you, you cut your self." *Yudas bilong em i-hankamapim em yet.* "Judas hanged himself."

Certainly, *yet* can also be used in the function of an intensifier, as the examples below show:

(240) Q: Man i   salim meri? A: No, meri **yet** i   salim man.
Q: Man PM send woman A: NEG woman FOC PM send man

Q: The man sent the woman? A: No, it was the woman who sent the man. (ZC 02377)

Q: Em dispela mama i   makim ol? A: No, papa **yet** i
Q: 3SG DT   mother PM make   3PL A: NEG father FOC PM
makim, em   mekim (t)oktok olsem long   ol.
make   3SG make talk   same PREP 3PL

Q: She, this mother did them? A: No, it was the father who did, he told them these things. (ZC 007413)

Naturally, the fact that *yet* can be both used as an intensifier and a reflexive pronoun does not prove that it evolved from an intensifier to a reflexive pronoun and not vice-versa, or with other steps involved, and even if that path is correct, how it came to be used as an intensifier in the first place. The last point, however, was convincingly detailed by Sankoff (1993). Regarding the origin of *yet* as an intensifier, she writes (1993, 138):

> Speakers of Austronesian languages in the contact situation heard English *yet*, used mainly as a negative polarity item, but also sometimes as the positive 'still'. Tolai speakers identified it with their own *iat*, which is used as an intensifier but is also associated with the meaning 'still'. *Iat*, however, has focus-marking functions that English *yet* has not. *Yet* was pressed into service as a focus marker in TP, and also, over time, drastically reduced its function as a negative polarity item. Thus today its distribution in TP is split between adverbial intensification, and the focusing of nominals, much the way *iat* works in Tolai, and the way other post-verbal emphatics, focusers, reflexives and limiters do in Austronesian languages. It is not surprising that in the area of focus, with its great importance in rhetoric, the speakers of TP have shaped their language to fit the patterns their ancestors have used from *bipo yet.*

She reports the following numbers for the various functions of *yet* across her sample of speakers (1993, 131):

	Chavi	Mr. Mo.	Adults	Children	Total
temporal - neg. 'not yet'	11	8	-	1	20
temporal - pos. still	3	10	8	6	27
intensifier w. adj./adv.	-	3	3	-	6
focus particle w. pronoun	4	8	13	4	29
Total	18	29	24	11	82

Table 6.31: Instances of *yet* in various functions in Sankoff (1993)

Table 6.32 shows the same functions across a sample of 249 instances of *yet* in the Z'graggen corpus, using Sankoff's criteria and categorization. The five remaining instances of *yet* were as a focus particle with something other than a pronoun. As we can see from the table above, the positive temporal adverbial function is by far the most common in the Z'graggen corpus. As Sankoff suspected, the negative function 'not yet' is by far less common - 16 to 149 occurrences.

temporal - neg. 'not yet'	16
temporal - pos. still	149
intensifier w. adj./adv.	15
focus particle w. pronoun	64
Total	244

Table 6.32: Instances of *yet* in various functions in the Z'graggen corpus

Unlike Sankoff's samples, the Z'graggen corpus contains some examples of *yet* as an intensifier with adjectives and adverbials. Two of these are reproduced below:

(241) Na   bipo   **yet** Niu Gini   i   no   olsem nau yumi
       CONJ ADV-T INT New Guinea PM NEG same now 1PL-INCL
       i   stap long   em.
       PM be   PREP-L 3SG
       And way before, New Guinea was not the same, now we are here. (ZC 00253)

(242) I   go insait na   i   skruim long   dispela gen   na   i   go
       PM go inside CONJ PM insert   PREP PRN   again CONJ PM go
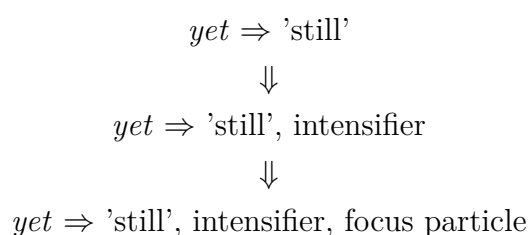
moa **yet**, i   go moa   yet.
ADV INT PM go ADV INT

He went inside and inserted into it again and went even further, even
further. (ZC 04077)

Notably, there are also instances in which *yet* is ambiguous and it cannot be
inferred without context whether it is a temporal adverbial itself or intensifies
a preceding temporal adverbial:

(243)   Na   wanpela meri,   oh i   stap longwe   **yet**.
        CONJ NUM   woman oh PM be   far-away INT/ADV-T

        And one woman, oh she was very far away/far away still. (ZC 00671)

Such ambiguity would have further increased the likelihood of an expansion
from a temporal adverbial to an intensifier. We can therefore conclude that, as
Sankoff describes, *yet* has "undergone a shift away from what was apparently
its most general use in the 1920s, a temporal negative polarity item, and to
have undergone the following evolution" (1993, 132):

$$yet \Rightarrow \text{'still'}$$
$$\Downarrow$$
$$yet \Rightarrow \text{'still', intensifier}$$
$$\Downarrow$$
$$yet \Rightarrow \text{'still', intensifier, focus particle}$$

However, as evolution is wont to, it did not stop at that point. Rather, as
mentioned above, it expanded further from focus particle on pronouns to the
reflexive pronoun. This does not require much effort either syntactically or
conceptually. Consider that in the same way the example above was ambiguous
between an intensifying use and a temporal adverbial use of *yet*, the examples
244 through 246 below are syntactically ambiguous between an intensifying
use and a reflexive use[114]:
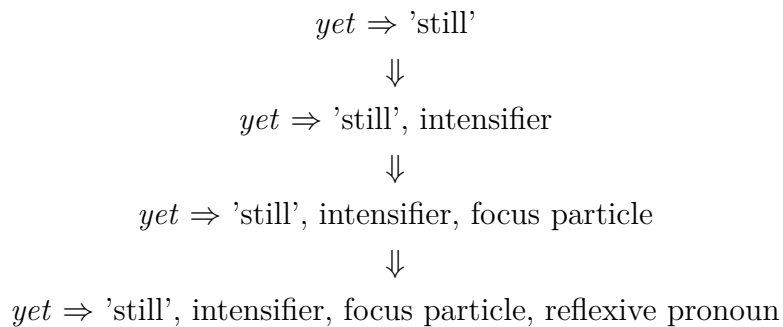
(244)   Em redim   kaikai pinis   orait   em bilaisim em   **yet**
        3SG prepare food   COMPL alright 3SG paint   REFL/INT now
        nau.

        He finished preparing the food, alright, then he painted himself. (ZC
        03788)

---

[114]Note that they are ambiguous both in Tok Pisin and in English.

(245)   Orait   bai   mi   raunim mi   **yet**   bai   mi   tromoi   tel
           Alright FUT 1SG chase   1SG REFL/INT FUT 1SG throw-out tail
           bilong mi   i   go daun long   yu.
           PREP 1SG PM go down PREP-D 2SG

           Alright, I will chase myself, will throw out my tail and go down to you.
           (ZC 04624)

(246)   Olsem   na   meri   bilong   em tokim em, Oh mi   ting
           All-the-same CONJ woman PREP-P 3SG tell   3SG Oh 1SG think
           yu   man nogut, yu   gutpela man na   yu   haitim yu
           2SG man bad   2SG good   man CONJ 2SG hide   2SG
           yet,   yu   laik   giamanim mi.   (ZC 00999)
           REFL/INT 2SG want lie   1SG

           All the same and his wife said to him, Oh I think you're a bad man,
           you're a good man and you hide yourself, you want to lie to me.

The analysis of *yet* as a reflexive pronoun is hindered by the fact that it is quite rare in the available data. For *mi Vtr *im mi yet*, there are only four occurrences, three for the same construction with *yu*, two for *mipela* and six for *ol*. Occurrences with *em* are difficult too analyse due to the fact that the pronouns frequently have different referents. Still, it seems highly probable that we can complement the evolution drawn above by a further step:

$$yet \Rightarrow \text{'still'}$$
$$\Downarrow$$
$$yet \Rightarrow \text{'still', intensifier}$$
$$\Downarrow$$
$$yet \Rightarrow \text{'still', intensifier, focus particle}$$
$$\Downarrow$$
$$yet \Rightarrow \text{'still', intensifier, focus particle, reflexive pronoun}$$

It bears mentioning that *yet* is, in its evolution, unique among the features surveyed in this analysis. First, its main development occurred quite a bit later than the others. As Sankoff points out, the shift from the temporal negative item occurs after the 1920s. Secondly, it is among very few features for which significant substrate influence can be assumed. Note, however, that this substrate influence once again came from Tolai, an Austronesian language, and not an Oceanic substrate.

7 Discussion

In general, I believe we can draw the following conclusions from the data analysed above:

1. Tok Pisin exhibits some unique structures and grammatical items/patterns, but not necessarily unique functions when compared to other English-lexified contact languages.
2. These items and patterns arose, for the most part, not from direct borrowing from either super- or substrate languages, but were innovated from material existent in expanding Tok Pisin.
3. The timing of the origin of many of these innovations is highly relevant as it determines both the extralinguistic factors of the emergence situation as well as the number and structure of existing structures within the language.
4. Many of the most intriguing grammaticalization phenomena are related to increasing syntactic complexity, whereas lexical and morphological development as well as tense, aspect and mood marking has followed more "traditional" paths.
5. When analysing the grammaticalization path of an individual structure, it is relevant to analyze competing structures which have arisen through competing or converging grammaticalization paths as well.
6. Variation and competition between two or more different structural items and patterns can be the result of varying grammaticalization paths and origins.
7. Emergence of individual variants has to be investigated in full context (see abilitative complex).
8. Both semantic and syntactic reanalysis are fundamental in the emergence of the structures analysed.
9. Innovation as analyzed happens mostly through reanalysis rather than through active innovation in which the agency of the speakers plays a major role.

10. Most of the syntactic expansion happened through reanalysis of existing material. However, none of this material was taken from the Oceanic substrates. A strong substrate influence on syntactic expansion as suggested by Keesing could not be found, given that all phenomena could be accounted for by language-internal development. This strenghtens Mühlhäusler's model and casts further doubt on very early stabilization and relexification as suggested by Keesing.

7.1 Form and function

There is no question that Tok Pisin displays unusual or, at the very least, uncommon structures. When compared to other English-lexified contact languages, it frequently deviates from a common pattern in the form it chooses to express a certain grammatical function. Among such examples as explored above were *wantaim* as a comitative and instrumental preposition (as well as a later coordinative nominal conjunction). Both in terms of its concrete morphological structure and as a derivation of English *one time*, it does not appear in any other English-lexified contact language or any other language, for that matter. But still, in its individual functions, it is not innovative. There is nothing unusual about a language having a comitative preposition, or an instrumental preposition, and neither is there about a language employing the same item or construction for both functions. One does not need look far to find them - English *with* and Bislama/Solomon Islands Pijin *wetem* are identical in function. The only difference, then, is in the process in which the specific form and its specific function came to aligned. This process is nothing unusual either. It is simple semantic reanalysis on account of specific properties of the original construction *one time along*, as shown in chapter 6.2 above. In addition, *wantaim* has - on a well-documented grammaticalization path - also expanded to fulfill the role of a nominal conjunction.

Similar observations can be made for *tasol*, which also displays no fundamentally unique or unusual function as an adversative conjunction, emphatic particle and exclusive particle. Its individual functions arise from each other via processes of language change that have been observed in other languages, including those where language contact has played no significant role. Again, the form itself has not been adopted directly as a form-function pairing from any of the contributing languages. The innovation is not in the form itself - which has been adopted from English - nor in its functions, all of which are

observable in other languages, but in the recombination of both the form-to-function mapping and the individual functions that the form serves. Tok Pisin, or rather, its speakers, have innovated a novel form-function pairing that is not observed in any other English-lexified contact languages, including its sister dialects of Bislama and Solomon Islands Pijin.

A third example in this regard comes from various TMA markers. Here, we see some parallels with Bislama, for instance in the common progressive marker *i stap*, although, again, Tok Pisin exhibits an additional progressive marker in *wok long*. Yet in terms of function, they are not unique to Tok Pisin: there are plenty of languages which offer a grammatical means of marking the progressive, including the *-ing* form in English or the emerging *am +* V construction in German (which may or may not eventually fully grammaticalize). Neither are other markers such as *inap* unique in their function, even though they are an unusual form-function pairing.

Tok Pisin thus can be said to have unusual form-function pairings. Such pairings can, theoretically, be unusual in three possible ways:

1. A form existing in other languages exhibits a function not present in other languages (existing form, entirely new function)
2. A form not existing in other languages exhibits a function not present in other languages (new form, entirely new function)
3. A form not existing in other languages exhibits a function present in other languages (new form, existing function)
4. A form existing in other languages exhibits a function it does not fulfil in other languages (existing form, existing function that is expressed by other forms in source language)

To clarify, the first option would be that a form exists in another language, maybe even one of the source languages, but exhibits a function that is not present in any other language, or, at least, very few languages. This would mean that speakers of Tok Pisin would be able to grammatically encode concepts that speakers of most other languages are unable to encode. This is not the case. Functionally, Tok Pisin shares its inventory with many other languages. It has no rare grammatical moods such as the mirative in Turkish or Korean (deLancey 1997)[1] or other typologically rare features such as case stacking or pluractionality. Hence, the second option is also not present in Tok Pisin.

Indeed, the 'unusual' factor in Tok Pisin is almost exclusively tied to the

---

[1]For those curious and unfamiliar with the mirative, it is a grammatical means of expressing the speaker's surprise or unpreparedness.

third and fourth options. Forms are innovated to convey concepts that have other form-function pairings in source languages, but can be expressed in them. Depending on the construction in question, the form-function pairings in Tok Pisin either consist of a new form expressing an existing function, such as in the case of *wantaim*, *tasol* or *i stap* (option three), or of a form borrowed from a source language but assigned a new form in the development of Tok Pisin, as has been the case for *-im*, *-pela* or *bin*.

The second point of note in this regard is that the constructions in and general structures of Tok Pisin emerged not through processes inherently different than those other contact languages, or, in fact, other "normal" languages emerged through. Semantic and syntactic reanalysis as well as the general process of grammaticalization are inherent, central parts of language change as described in countless other instances. It is not the processes of change that differ here, but the input and the environmental factors - language-internal and language-external - that have determined "unusual" outcomes. This strengthens arguments in favour of seeing pidginization and creolization not as radically different processes of language change, but processes of language change working under different circumstances and with different materials. The processes may be sped up due to a need for rapid establishment of certain structures. The material may be more diverse than that available in general language contact. There may be more variants to select from, and inter- and intraindividual variation may be greater. But the processes run along the same lines. The material is eventually levelled. Variants which offer greater advantage in communicative purposes or which receive boosts from analogous constructions or some other kind of evolutionary advantage are selected, whereas others become more rare and are eventually removed from both the common feature pool and the one of individual speakers. The processes that govern these changes, however, remain the same.

Note that whether the need for the function was triggered by contact (see contact-induced grammaticalization and the notions of linguistic supply and demand in chapter 2.6) makes no difference at this point: the form arose through internal innovation in Tok Pisin, which brings us to the next point.

7.2 The origin of innovation

The second conclusion to be drawn from the analysis above is that Tok Pisin exhibits many constructions that were innovated within the emerging language system, and not directly borrowed from any source languages. Exam-

ples abound. We have to differentiate, however, between items such as *-im* and *-pela*, which arose solely through syntactic reanalysis early on in the development of Tok Pisin and the other Neomelanesian dialects, and items and patterns that were part of the later, mostly syntactic expansion, such as the connectives, relativizers and complementizers. While the former have arisen mostly before the linguistic system of Tok Pisin split from the other Neomelanesian dialects, this does not invalidate the point I am trying to make: the form-function pairings were not borrowed directly from any source language, be it a super- or substrate. English, German and Papuan structures triggered a functional demand in the language, leading to contact-induced grammaticalization. The incomplete acquisition of any superstrate led to syntactic reanalysis of certain items, such as *-im* or *-pela*.

In the end, however, neither of the two groups were directly adopted from English as a superstrate language. The former were results of syntactic reanalysis of English lexemes. One could argue that they were the result of failed or imperfect transmission. However, this would imply that the primary intention of the speakers who innovated these forms was to *replicate*, when in reality it would have been to *communicate*. As such, these forms do not present a failed attempt at replication, but a successful attempt at communication drawing on the available resources to fulfil communicative, cultural and societal needs in a creative way.   The same is true of the second group, which was reanalysed or otherwise innovated when the input of English was no longer present. It therefore relies less on the syntactic position and function that the items had in English, and more on the position and function they had in the already emergent contact language. Development among these items ran parallel to that of the original English items, but either no or not all functions were directly borrowed. The same is true, of course, for the German superstrate, which left even fewer structural traces on the system of Tok Pisin. Substrates, in turn, seem to have influenced some grammaticalization processes, such as that of *yet* as a relativizer. But even in these cases, the process was not as simple as the substrates providing structural pattern for Tok Pisin to model their own intensifier after. Rather, it was the concurrence of English *yet* and Austronesian *iat* that led to *yet* adopting additional functions modelled after the substrate. Again, in these instances, speakers drew on the available material and its structural constraints in a creative manner to introduce new functional items.

Most of the examples surveyed above can be motivated, in terms of their form-function pairing, by language-internal change. They do not require, for the most part, structural influence from either super- or substrates, only the

intention of speakers to communicate and innovate by drawing on existing structures. *Na*, for instance, originated from the position and function that *nau* had in emergent Tok Pisin, not the position and function that adverbial *now* had in the English input that preceded it. If it had not been used as a discourse marker in narratives as much as it was, it is unlikely that it would have developed into the coordinative conjunction - or maybe it would have, but certainly by another grammaticalization path. And had it had a similar position and original function in English, then English speakers might be able to use *now* as a conjunction.

The same argument can be made for the interplay between the anteriority and past markers *bin* and *pinis*. As described above, at some point speakers of Tok Pisin had access to two past markers in these two forms. However, due to the inherent syntactical and functional structure of the language system at the time, a competition between these two forms for past marking never played out. Instead, they each specialized into another function, one becoming a marker of anteriority, while the other retained its status as a completive marker. Such a resolution was only possible in these specific circumstances, and may have played out differently in any other language whose speakers had access to two variants of the past marker. Thus, even in a case in which the surface forms of two competing variants were taken directly from English, the key parts of their functional development played out entirely within the emergent linguistic system of Tok Pisin.

The grammaticalization of *ia* was similarly dependent on the properties of Tok Pisin as a whole. Consider that not only was there a close semantic analogy between the use of *ia* as a demonstrative and a spatial adverb, but that there was also a close semantic and syntactic analogy between *ia* as a demonstrative and *ia* as a relative marker. This led to the fact that sentences with *ia* in such a position were ambiguous, which in turn led to the reanalysis of *ia* as a relative marker. Again, such developments were reinforced by substrate influence from Austronesian languages, in which there were similar structures. The mixture of ambiguous placement, substrate influence and reanalysis meant that *ia* developed in a certain direction within the emergent system of Tok Pisin. As other languages would have a different set of such circumstances, we do not see the same development in them. Again, the specific structures and the specific input necessary for *ia* to become a relativizer combined to pave the way for a language-internal innovation that was spurred on by external factors and the contact situation, but ultimately did not constitute a borrowing from one of the source languages.

Thus, to understand how constructions such as these came to fulfill the

functions that *na* and other items with similar origin have in Tok Pisin, it is necessary to determine both the form and function of the original borrowing - in this case, *nau* - and the period in which it was borrowed, including how it functioned in the superstrate as well as the internal development and grammaticalization it was then subjected to in Tok Pisin, with or without possible influence from English. In any case, I believe that more research into structural innovation in contact languages is worthwhile. If, as Roberge (2009) and others claim, pidgins can give us a possible window on language evolution, it stands to reason that those instances in which the emerging structures are, to some extent, innovated instead of drawn directly from the super- and substrates, would be the ones that are most informative regarding the process of language evolution. After all, it would be those instances that would have to rely on the processes of linguistic innovation the strongest, leaving aside, for the moment, those instances in which linguistic innovation happens without **any** previous input. Since there are no cases of the latter, it is these cases of emerging contact languages that provide the best opportunity for studying such processes: the structures are innovated within a system that has not yet reached full functionality or full stability and the speakers are motivated to innovation by a need for successful communication that cannot be fulfilled through falling back on another language, given that they lack another common lingua franca.

7.3 The timing of innovation

An important factor determining the inception, diffusion and eventual outcome of any innovation - be it entirely internal, externally motivated but internally structured, or just a direct borrowing - is the time at which it occurs. This affects the process in two significant ways: first, the timing determines the extralinguistic factors that come into play, which in turn determine the available input, the prestige of various variants, the number of speakers in the speech community and many more. Secondly, emergence later on the timeline of development means there are more existing structures within the language taking possible effect on the innovation, whereas innovations that happen earlier during the emergence of a language have less possible interference effects, but also fewer analogous processes and structures to draw on.

Let us examine each of these points in detail. In regards to point three, the timing of these innovations, consider the following. As we have seen, some of the innovations covered above were introduced and/or asserted themselves at a

point in time when the input of the English superstrate was weakened, having shifted from an L1 to an L2 input, or completely absent. One prime example of these innovations is, of course, the plural marker *ol*. With its development roughly dated to the expansion stage between 1880 and 1914, it falls squarely into the period of time where English input would have been limited due to the annexation of north-eastern New Guinea by the German Empire. As such, part of the reason why plural marking is not directly borrowed from English is that the timing of functional demand in the emergent linguistic system of Tok Pisin and formal supply from English did not link up.

This spot, however, was not filled by German either, whether for reasons of incompatibility with the already emerging contact variety or due to similarly inconvenient timing. Instead, speakers exploited language-internal material to innovate strategies and patterns that allowed for expansion. The same is true for *wantaim*, which emerged as a preposition during the same period. Again, the input of English was present earlier, and if it still had been at the time speakers of Tok Pisin were in need of a dedicated comitative and instrumental preposition, we might see a form of *with* in today's Tok Pisin, similar to the forms we see in neighbouring Bislama and Solomon Islands Pijin.

The timing for the adoption of constructions in earlier stages of Tok Pisin and preceding pidgin stages is also relevant. In general, the specific point in time at which a construction is introduced and diffused into a language system also determines the extralinguistic factors that affect its development. This is especially crucial in languages shaped largely through language contact, since it determines the available input and the socioeconomic relations between the speakers of the various input languages. Thus, those innovations adopted through possible "interrupted" or "failed" transmission or incomplete language acquisition are only possible during periods in which one of the source languages is still the dominant target language for speakers. In the development history of Tok Pisin, this includes mostly early reanalysed forms such as *-im* or *-pela*. They were only originally available as forms due to the input of L1 English being present.

Another way in which timing is important to the innovation in question is that it affects which other structures and complexities are already present in the language. For instance, the use of object fronting for focusing nonagent participants is only possible due to a set standard word order. In this regard, timing is of course not randomly determined: the needs of the speakers grow increasingly more complex in terms of structure as the emergent language becomes used in more and more circumstances and needs to be able to express more and more complex concepts. It is quite likely, for instance, that a rela-

tivizer or a complementizer will always develop later in the diachronic development of a contact language than will, for instance, TMA marking strategies. In other words: as complexity of the linguistic system as a whole increases, so do possible influences of existing structures on innovative structures.

The fact that timing of innovations is crucial is also reflected in the fact that some of the innovative constructions surveyed above are now in the process of being challenged, at least, or replaced, at most, by renewed borrowing from English as contact with English becomes more prevalent again, especially in the urban dialects of modern Tok Pisin. An example is *bat*, which was specifically not borrowed into Tok Pisin at an earlier time as described above. This lack of a borrowing allowed for the innovation and expansion of *tasol* as the adversative conjunction. Now, *bat* and *tasol* are in competition for the same function as an adversative conjunction. As Dutton (1999, 30) and Smith (2002b, 187) describe, the use of *bat* in this function is increasing especially among younger speakers. Similarly, Mühlhäusler (2003, 12) notes that the "autonomy and consistency of Tok Pisin are threatened by the renewed contact with English", noting that there is increased use of English pronunciation, such as *got* instead of *gat*, an increase of traditional vocabulary such as *long* being replaced by English equivalents, such as *on*, and a loss of grammatical distinctions such as the blurring between inclusive and exclusive third person plural pronouns due to the introduction of *wi*. The latter can be used as either an inclusive or an exclusive third person plural pronoun in modern, urban varieties of Tok Pisin.

The replacement of such innovations by renewed borrowing indicates that a functional demand is not the only driving factor behind language change. All the English replacements listed in the paragraph above fulfill functions for which the speakers of Tok Pisin have already innovated other forms - yet they are still being replaced. Competition and selection do not stop once a dominant form has been established - neither in biology, nor in linguistics.

Furthermore, the importance of timing illustrates another indication that there is no factor inherent to the language structure of Tok Pisin which would make it more susceptible to innovation and creativity - it is merely the timing of the functional need which determines the available and innovated forms for a specific function, both in terms of the extralinguistic environment and in the existing complexity of the linguistic system as a whole, with the possible interferences that it carries.

7.4 Innovation in syntax

As for the fourth point, the difference in the grammaticalization paths for various areas of Tok Pisin's grammar, it is evident that many of the structurally divergent features in Tok Pisin are related to increasing syntactic complexity. Lexical and morphological development is far more closely aligned with the other contact languages surveyed. That is to say, Tok Pisin shares a great deal of its vocabulary and its morphological inventory with Bislama and Solomon Islands Pijin. The interesting differences are mostly found in those features which emerged later and therefore under circumstances of increased linguistic complexity. This is largely a function of the aforementioned point on timing: functional need for tense-aspect-mood markers, for instance, seems to have predated the need for syntactic functions such as relativizers and complementizers. As such, the contact with English would have been more intimate at the time that they were introduced. Thus, the motivation behind the adoption of certain TMA markers is semantically far more transparent than for some of the other features.

For instance, the early syntactically reanalyzed *-pela* and *-im* can be counted as morphological innovations. They require less structural foundation in the linguistic system than a relativizing construction. While *-im* requires only a base class of verbs to be productive, *ia* requires several syntactical relations to be in place in order for it to fulfil its function. This is, of course, a reflection of the complexity of the underlying concepts. Arguably, the cognitive concept behind transitivity is less complex than the one behind relativization, which requires two separate events or states and at least one entity, i.e. an agent, the event or state described by the main clause's predicate, and the event or state described by the relative clause's predicate.

Similarly, many of the TMA markers are easily directly semantically reanalyzed, whereas other items, such as *yet* or *ia*, take a less direct grammaticalization path. As such, there are some areas of Tok Pisin's grammar which are significantly less transparent than others. Most speakers of English would not have difficulties parsing much of its nominal and verbal system, whereas the syntactically more complex features would be a considerable challenge.

In general, one could say that there is a scale from less disruptive to more disruptive that innovations can fall on, with lexical innovations being the least disruptive to the linguistic system as a whole, whereas complex syntactical structures are most disruptive. Morphological innovations lie somewhere in between. For instance, innovating a new lexeme for a concept hitherto unrepresented in the vocabulary of any given language will not, at least not

immediately, affect the linguistic system as a whole. A new noun or verb can be integrated into a language without requiring any additional changes to its structure - at most, some phonological adaptation will be necessary. This is of course assuming that a nominal word class existed in the language at that point. The new lexeme (at first) affects nothing but its own introduction and the meaning range it can be used for.

The picture changes for morphological innovations. Let us take the transitivity marker *-im* as an example. Any such marker affects not only its own structural form, but, should it become productive, enables changes to the entire morphological and functional category it represents. Through its inclusion, the linguistic system gained a new level, or node, depending on how one visualizes an increasingly complex system. This node in turn affects other nodes: first of all, the verbal inventory, which interacts with the new marker to derive new word forms. Secondly, other syntactical innovations, which relies on the possibility of marking transitivity, are now possible. The system as a whole has been affected to some extent.

This pattern is doubly true for syntactical developments. Rather than affecting the function of single lexemes or classes of lexemes, they enable complex concepts such as relativization. In order to do so, as mentioned above, they touch on many other nodes, including the ones responsible for linguistically encoding such concepts as agentivity, states, events and others. Since these are usually encoded by different forms, the relativization node in the linguistic system needs to be connected to and in turn affect all of these. Conceptual efforts such as these are only possible by affecting wide arrays of the system. Not only are the motivation for *yet* as the reflexive pronoun or the bracketing of *ia* as a relative marker less apparent, but the structural level of the linguistic system at which they come into play is also far more complex than that of an adjectival suffix.

All of these are only tendencies, and only limited to the processes of emergence detailed above. I do not mean to suggest that Tok Pisin developed by first exclusively innovating "simple" or purely lexical structures, then morphological structures, and finally syntactical structures. By necessity, there has to have been some overlap, although the tendency must have been for more syntactically complex structures to emerge last. Neither do I want to suggest that the processes behind the inception, diffusion or grammaticalization of structures is inherently different between morphological and syntactical innovations, or for innovations dating earlier or later in the emergence of a contact language.

Similar to how these processes do not inherently differ between languages

for which contact plays a large role, and for those in which it played a smaller role, grammaticalization processes do not change fundamentally whether they are applied to more or less complex structures. However, they too grow more complex with more factors of the system affecting their individual subprocesses. A more complex system, for instance, offers multiple avenues of grammaticalizing a certain function. In other words, more grammaticalization paths open up. Consider, for instance, that Tok Pisin has several ways of expressing the relativizer. This is only possible because there are several constructions which originally served another function on said grammaticalization path. *Ia* bracketing, for instance, tread this path from 'here' to the demonstrative to the relative, whereas *husait* and *we* tread it from 'who' and 'where' to wh-question to the relative. Similar arguments could be made for *wantaim* and *na* or *wok long* and *i stap*. In this regard, complexity begets variation, and variation begets competition.

7.5 Insights from competition

Points five and six are closely interwoven, so I will deal with them both under this section. The fifth point stresses that divergent and convergent grammaticalization paths can lead to competition between several forms for the same function within a language. This is evident in many of the examples above, but just to reiterate, here are three examples of this phenomenon. For the abilitative complex:
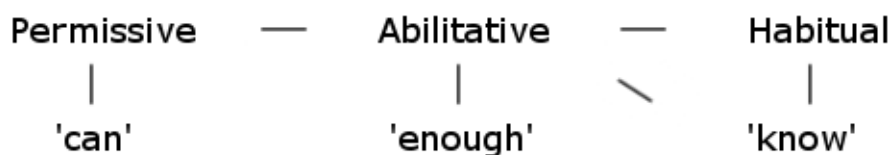

Figure 7.1: Grammaticalization paths for the abilitative

There are three variants competing for the function of the abilitative. There is *ken*, which has emerged through the permissive path. There is *inap*, which has emerged directly as an abilitative. And there is *save*, which, in addition to being a habitual markers, also has directly emerged as an abilitative.

The relative markers paint a similar picture: *husait* and *we* have evolved from the path of wh-question, whereas the system of *ia* bracketing has emerged

```
'who'                                    'here'
  |                                        |
wh-question  —  relative  —  demonstrative
  |
'where'
```
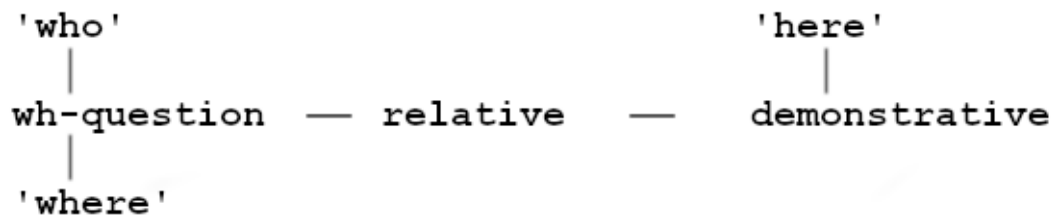
Figure 7.2: Grammaticalization paths for the relative

through the demonstrative path. Both have extended to the relative from their original function and are now competing for this additional function.

```
progressive  —  continuous  —  existential
     |                              |
    'do'                        locative
                                    |
                               'stay, remain'
                                    |
                                  'stop'
```
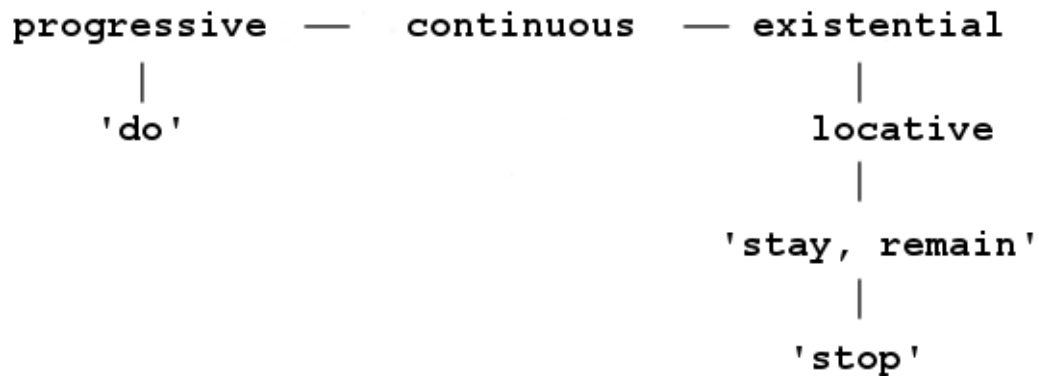
Figure 7.3: Grammaticalization paths for the progressive

In the map for the progressive, as has also been described above, there is an imbalance in the starting points for *wok long* starting from 'do' and *i stap* starting from 'stop'. However, as *i stap* is an older variant than *wok long* for the continuous, they are now in competition for this function. *Wok long* has, as described above, grammaticalized there from the progressive, whereas *i stap* took a longer path through the meaning of 'stay, remain' to the locative and the existential. In the end, however, these paths are converging - one is almost tempted to use a tired expression such as "all roads lead to Rome" (which would be false in any case, since not all grammaticalization paths end up at the same function).

In regards to the sixth argument, which follows from the fifth, there is a synchronic and a diachronic aspect. For the former, analyzing competing constructions and their grammaticalization paths can give insights into variation

in the current use of the language. For instance, the fact that as a conjunction, *wantaim* arose from the comitative preposition and therefore has a closer link to noun phrases, while the competing coordinative conjunction *na* arose from a verbal conjunction and therefore has a closer connection to verb phrases can help explain the fact that we see the former occurring more often as a nominal conjunction than the latter. In a similar vein, the fact that *we* occurs primarily (or exclusively) as a complementizer within a relative clause, while other complementizers such as *olsem* can occur outside of the scope of a relative clause, tells us both something about the origin of *we* as a complementizer and about the origin of the other complementizers. Similar statements can be made for the preference of *husait* for animate subjects when compared to *we*, or for the prevalence of *ken* as a permissive marker.

On the other hand, synchronic variation can help hint at the reverse side of the coin: strong patterns in systemic variation suggest different grammaticalization paths. In any case, the point I am trying to make is that variation among these constructions is by no means random. It is determined by the origin and functional history of the competing variants.

It would be interesting to see, however, whether there are cases in which the variation is truly random, with the individual variants being used in free variation rather than in complementary variation (or rather, somewhere on the scale between these two options that is closer to free variation than observed for these constructions). That, in turn, could mean two things:

1. both constructions emerged from the same grammaticalization path and therefore have the same functional scope
2. the grammaticalization process has taken place long ago and the two variants are detached from any functional associations with their 'older' functions

The first option, I would contend, is rather unlikely, given that in order to be equally likely to diffuse, the variants would also have to have a similar form. The second option, however, is not unlikely at all: if no variant is selected and successfully established as the preferred choice for all usage contexts, it may very well be that both variants continue in the feature pools of speakers without any meaningful difference. In reverse, it would be equally possible for variants that originally do not differ significantly to acquire differences in usage over time.

If it is not random which variants come into competition with each other, neither is it really random which variant is eventually selected as the dominant one. As detailed in chapter 2.3.2 above, there are various factors which

can make a variant more accessible, i.e. closer to the surface of the feature pool, or 'drag it down' and make it less accessible. Advantageous factors include analogous constructions within the same or within a source language, as we have seen, for instance, with the establishing of *yet* as a reflexive pronoun. In other cases, such as the emergence of the plural marker *ol*, the use of reduplication in derivational processes have disadvantaged its adoption as a morphological means of plural marking. For ongoing competitions, the discourse marking function of *na* in narratives give it an advantage in its function as a predicative conjunction, whereas the comitative function of *wantaim* give it an advantage in nominal conjunction. The spatial connotation of *we* makes it a preferable choice as a relativizer for spatial nouns such as *ples*. The list goes on and on.

Finally, synchronic variation is the starting point for diachronic change, or at the very least, one possible starting point. Variation may be present in the synchronic feature pool of a speaker, but the competition that plays out between the variants over time can lead to lasting change.

The final point I would like to make in this regard is one that might seem trivial to some readers: do not look at grammatical or syntactical phenomena in isolation. This recalls the sorting of TMA markers in the analysis chapter above into, among others, an abilitative complex and an inceptive complex. No grammaticalization or other process of language change happens in isolation. It is affected by other, competing variants, by existing structures and by extralinguistic factors. In order to adequately explain the origin and current function of a construction in question, we have to look at what other functions it relates to and how, which variants it is in competition with, which other variants it already has replaced and what form it has taken as well as to why and how functional demand is present at that stage of development.

7.6 Semantic and syntactic reanalysis

Point number seven details that when it comes to the processes by which many of the innovations above were introduced into the language, maybe the most important was reanalysis. I have summarized the primarily responsible processes in the table below:

form	main function(s)	primary process
wantaim	comitative/instrumental	semantic reanalysis
-pela	adjectival suffix	syntactic reanalysis
-im	transitive suffix	syntactic reanalysis

i	predicate marker	syntactic reanalysis
nau	coordinate conjunction	grammaticalization
o	disjunctive conjunction	grammaticalization
tasol	adversatice conjunction	grammaticalization
maski	concessive, prohibitive	syntactic reanalysis
bin	anterior marker	syntactic reanalysis
pinis	completive marker	semantic reanalysis
i stap	continuous marker	semantic reanalysis
wok long	continuous marker	semantic reanalysis
kamap	inchoative, ingressive	semantic reanalysis
kirap	inchoative, ingressive	semantic reanalysis
save	abilitative, habitual	semantic reanalysis
inap	abilitative	semantic reanalysis
ken	permissive	semantic reanalysis
traim	attemptative	semantic reanalysis
-	passive	structural innovation
ia	relativizer	semantic reanalysis
we	relativizer	semantic reanalysis
husait	relativizer	semantic reanalysis
long	complementizer	syntactic reanalysis
olsem	complementizer	syntactic reanalysis
na	complementizer	syntactic reanalysis
we	complementizer	syntactic reanalysis
yet	reflexive	substrate influence

Table 7.1: Primary innovation processes for the constructions surveyed

From this, three patterns emerge. First, groups of constructions similar in function tend to arise through the same process. For instance, the three conjunctions all arose through semantic reanalysis, both progressive markers (and, in fact, most of the TMA markers) arose through semantic reanalysis, all relativizers arose through semantic reanalysis, while all complementizers arose through syntactic reanalysis. Another group are the earliest constructions surveyed, the adjectival suffix *-pela*, the transitive suffix *-im* and the predicate marker *i*, which all arose through syntactic reanalysis. This could mean several things. For one, it is possible that similar functions arise through similar processes. This is also tied in with the fact that, naturally, most of the forms competing for similar functions above arose during roughly the same developmental stage of Tok Pisin, given that that particular stage had a functional demand for the particular function. This would make both the available input and other extra- and intralinguistic factors similar for the constructions in question. In addition, a successful diffusion of one particular type of construction

influences the future success of other, similar constructions and developmental processes by virtue of analogy.

Secondly, it is hard to distinguish whether a construction arose solely through semantic or syntactic reanalysis, which is why the summary above should be taken as either semantic properties or syntactic position/function should be taken as primarily responsible for analysis, but not necessarily as solely responsible. Consider, for instance, that its syntactic position as a post-verbal marker affected *pinis* in its competition with *bin* and the eventual specialization of both forms. Similar observations can be made for *save* and *i stap* (see chapter 6.8.2). Additionally, syntactic reliances affected the reanalysis of *inap* as an abilitative (see chapter 6.8.4). I would therefore be wary in ascribing sole responsibility to either semantic or syntactical properties in some of these cases, though certainly one was the prime driving factor. For others, usually the ones in which syntactic reanalysis is dominant, that is not necessarily the case: we can consider *-im* and *-pela* to be instances of pure syntactic reanalysis.

Thirdly, in those cases in which there is competition between several forms for a single function, that competition is different for those constructions which arose primarily through semantic reanalysis versus those that arose primarily through syntactic reanalysis. While the latter more frequently occur in free distribution, the former often show a remaining pattern of complementary distribution, although rarely an obligatory one. As I have noted above, there is no strong contrast between the uses of *olsem* and *na* as complementizers, for instance, while there is stronger contrast between the uses of various TMA markers such as between *wok long* and *i stap*, between *kirap* and *kamap* or between *save, inap* and *ken*.

7.7 Innovation and agency

As mentioned in point number eight, syntactic and semantic reanalysis are the main driving processes of innovation in the Tok Pisin constructions surveyed above. Neither syntactic nor semantic reanalysis requires a high level of agency among the speakers of the language in question. As such, innovations with higher agency levels are, for the most part, confined to the lexis of Tok Pisin. It is here that we find the "creative" solutions to the problem of communicating a specific concept, such as *skru bilong lek* 'ankle', *gras bilong dok* 'dog's fur' or *as bilong diwai* 'tree stump'. For lexical constructions such as these, it is

likely that speakers consciously modeled them on the basis of either applying productive patterns to them or through circumlocution.

Syntactic reanalysis in particular seems to be a process that lacks agency on part of the speakers and hearers: they are neither responsible for consciously placing the item in question in said position, nor do they actively choose to interpret it in its new function. While this is less true for semantic reanalysis - for which there is a link in conceptual motivation - the fact that many of the semantically motivated grammaticalization paths are so well attested in other languages casts some doubt on the innovative potential here as well: if the inception process was highly agentive, we would likely see a lot more variation.

Relating back to the permutations of creativity in chapter 2.2.3, it is clear that for most of the constructions, they are not innovative in terms of semantic properties, nor are they in terms of their morphophonemic properties: they are existing constructions with existing meanings that, either by virtue of their semantic properties or through their syntactic position, get reanalyzed to serve a new syntactic function. This means that an innovative property is happening solely in the innovative level. While there is some adjustment over time on the morphophonemic level in the grammaticalization process, usually loss of material, it is irrelevant for the inception process.

Such observations mirror the discussion in chapter 2 above. It is the lexical level of a language that is most easily changed by a single speaker, and which has the least ripple effects towards the other levels. An agentive speaker is capable of introducing a new lexeme in a sentence without requiring her interlocutors to be aware of its function in advance in order to understand the structure of the sentence itself. Such a feat would not be possible with the syntactic or some of the TMA markers surveyed above. Secondly, we come back here to the notion of the invisible hand guiding the development of the language as a whole: the various factors raising the construction in a feature pool affect each individual speaker and therefore make the speech community as a whole more likely to use this construction. Thus, we see less variation the more advantageous one construction is, and more variation for those that have no clear competitive edges. If this development were more agentive, we would see either a lot more variation or a lot less, as speakers consciously choose one variant.

7.8 Innovation and the models

Point nine refers to Keesing's claim that much of the grammatical material in Tok Pisin stems from the Oceanic languages and was relexified with English surface structures. In other words, this would mean functional and formal supply in the first stages of adoption was supplied by Oceanic languages. Functional supply from English would have played a diminished role later on, with formal supply merely replacing already existent structures within Tok Pisin. This seems unlikely for several reasons due to the analysis conducted above.

First of all, almost none of the features surveyed displayed a strong substrate influence. Even in those cases where substrate features had a medium to large influence - such as in the case of *yet* - the origin of this substrate influence was not the Oceanic languages, but Tolai, and thus Austronesian. Furthermore, many of the phenomena analyzed can reasonably be assumed to have been adopted from English in a certain function directly, then reanalyzed within an expanding Tok Pisin, requiring neither an Oceanic origin, nor an Oceanic influence during their expansion stage. While this does not automatically preclude an Oceanic influence, it makes it more unlikely at least.

If, as Keesing suggests, Tok Pisin's structural development happen with some major setbacks, during which existing syntactical structure based on Oceanic substrates withered and was later relexified through Tolai or other languages, this would indicate that on a formal level, the corresponding constructions should match those found in Tolai (or said other languages). Yet as we have seen above, there are almost no grammatical elements among those constructions surveyed which show an Oceanic structure, but a Tolai form. For that matter, there are not even a significant number of Tolai morphemes or lexemes serving a grammatical function. Tolai itself seems to have mostly a substrate influence on Tok Pisin, rather than being a relexifier for earlier Oceanic structure.

It is of course possible that Oceanic substrate influence is more strongly present in features that are not part of this analysis, such as the pronominal system. If, however, Oceanic substratum influence was as pervasive as Keesing suggests, we would find traces of it across the linguistic board, not merely in specialized systems. Still, this is not to discard his notion that certain, early features of Pacific Pidgin did find their way into Tok Pisin, as is the case for *-im*, *-pela* and *i*.

Still, as I have shown above, there is sufficient linguistic motivation for the development of many of the expansion phase's syntactical and gramamtical structures without having to rely on an Oceanic substrate. Instead, construc-

tions such as the conjunctions, TMA markers, the complementizers or the relativizers can be seen as having developed within Tok Pisin, relying, for the most part, neither on direct borrowing nor on substrate reinforcement. This is once again not to say that such processes were not involved or that they were meaningless, but only that development can be inherently motivated and processed.

The timing of some of these expansions, as described above, is further indication for a weak or insignificant role of Oceanic substrate influence. Consider how some of the features described above are unique to Tok Pisin among the Neomelanesian dialects, while others are shared with both Bislama and Solomon Islands Pijin. It is mainly the features introduced and grammaticalized during the time period where Mühlhäusler locates the expansion phase which distinguish it both from its sister languages and its main lexifier. What implications does this have for the competing models of early stabilization (Keesing) and later stabilization and expansion (Mühlhäusler)? In order for the syntactic expansion of Tok Pisin to be successful, a previous stabilization stage is, of course, a necessity. However, by the assertions of both models, Tok Pisin would have been stabilized by the time the plantation workers were brought back to newly German-controlled Papua New Guinea. In Keesing's timeline, we would have seen a stabilization before the plantations, and stagnation of Tok Pisin's development during the plantation stage. In Mühlhäusler's timeline, the stabilization on the plantations is directly followed by the expansion of the post-plantation period. The fact that expansion occurred mostly after the plantation workers were brought back would, in my mind, speak in favour of Mühlhäusler's theory.

The possibility of Keesing's model and Mühlhäusler's model being combined into one model by resolving their direct points of contention remains. By removing the foundational importance of the Oceanic substrate from the former, a first step toward that purpose has been taken. The concession that stabilization is a stable process that affects all parts of a language and all parts of a speech community at the same time is a further necessary step in this direction. Indeed, Crowley (1990, 385) argues for abolishing the notion that one can neatly divide the history of Melanesian Pidgin into distinct stages. It is very possible, then, that stabilization occurred, to some extent, both before and during the plantation phase.

Note that the plantation phase itself did not lead to systematic expansion of the language. This only happened after 1884. One might take this as an indication that before the plantation phase, there had been no stabilization and the plantation phase itself must have been a necessary opportunity

to stabilize the pidgin before expansion could happen. This, however, is not necessarily true. As Romaine (1992a, 43) points out, there was little use for Pidgin English on the plantations in Queensland. With Melanesian workers having little access to English, they usually lived and worked in groups of similar origin, which means that there was little need to further develop the pidgin. There is also the fact that unlike Atlantic plantations, Pacific plantations were manned by voluntary workers, which would go back once their contracts were up, causing a constant renewal of the plantation population. This, along with the fact that the Melanesians never outnumbered the European population of the Pacific plantations in question, allowed Tok Pisin "to gain a considerable degree of structural complexity before creolization, and permitted substratum influence" (1992a, 43). The question of where and how Tok Pisin stabilized, then, should possibly be replaced with the questions of "did Tok Pisin ever stabilize as a whole" and "what aspect of Tok Pisin stabilized when".

8 An exceptional property

I would like to make three more points before closing this chaper and, indeed, the present work. First, even though I realize it is somewhat cliché to say this, a lot of work remains to be done. The analysis above is by no means all-encompassing or complete, in many regards, and I am aware of these limitations. It does not cover the entirety of the structure of Tok Pisin, for there are many lexical, grammatical and syntactical features that have not been analysed. It does not cover the entirety of the history of Tok Pisin, neither going all the way into the past (wherever the origin point would be), nor all the way to the present. Due to the limitations in available data, it covers, in its empirical analysis, only spoken varieties of Tok Pisin. In addition, many of the constructions surveyed arose in developmental stages that predate the data covered.

I do believe, though, that it is still a valuable number of steps towards the right direction. Much analysis remains to be done on additional data for the same phenomena. For written varieties, such an analysis might, for instance, be carried out on the Wantok corpus, although that too, hardly covers he entirety of Tok Pisin's history, nor its crucial periods of emergence. An analysis of competing and no longer competing variants for other structures of the language would be equally worthwhile, as would a more longitudinal study of the phenomena covered if sufficient historical data ever becomes available. A thorough analysis of a sufficient sample from the stabilization and expansion stages would be invaluable in this regard.

It would also be interesting to carry out similar analyses of the equivalent and of similar structures in the sister dialects of Bislama and Solomon Islands Pijin. In fact, I would deem an analysis based on competing variants worthwhile for all contact languages as far as the data allows, and possibly for any language.

In terms of interdisciplinary work, there might be insights to be gleaned from comparing frequent grammaticalization paths and sources with patterns in cognition and neurology. It is not inconceivable that those paths occurring with the highest frequencies mirror, in linguistic development, some cognitive

patterns or conceptual connections in our thinking (be the latter hard-wired or not).

My second, and second-to-last, point also regards interdisciplinary work, specifically the contribution of evolutionary concepts to the field of linguistics as a whole through the discipline of Evolutionary Linguistics. Of the numerous concepts covered above in chapter 2.3, I believe that three above all not only were beneficial to this analysis, but should be considered for further studies across the linguistic spectrum: the feature pool, competition and selection, and the notions of agency and the invisible hand.

The concept of the feature pool as adopted from biology's gene pool embodies two important ideas: first, that multiple variants can be available for the same function to the same speaker (and her idiolect). Secondly, that these variants are not necessarily equal as to their position within the pool and therefore the speakers' preferred access to the variants; some features gain benefits from analogous constructions or conceptual links, while others may benefit from their syntactical position or their morphophonemic appearance. The same features can, of course, also put a certain variant at a disadvantage. The inherent notion of the feature pool, then, is that variants coexist, and are not random.

This leads directly into the second point: variants are in competition and get selected according to their position within the feature pool, i.e. according to the various advantages and disadvantages that determine their position. Thus, not only is their inherent position not random, but the selection process is not random, either. These are, of course, two sides of the same coin: it is the same properties that determine the position within the feature pool and the selective advantages that a variant exhibits. The point, then, is that competition and selection do not occur randomly, but are determined partly by the inherent features of a variant and partly by how well they link up with the external circumstances variants get exposed to - similar to how certain genetic combination are only beneficial in certain circumstances.

Thirdly, the notion of the invisible hand explains the emergence of linguistic innovations and their diffusion. Agency over the inception of a concrete grammatical structure may lie with a singular speaker, but the diffusion is governed by network principles akin to that of the highway traffic analogy. Speakers can only directly affect those that they speak to, and if a certain variant that they introduce proves, in turn, beneficial to the communicative purposes of those speakers, it will spread across the speech community as a whole, as if guided by an unseen hand. Rather than the reflection of conscious decisions, however, this is a reflection of the fact that linguistic structures are

based on the same cognitive and communicative principles across the speech community. Therefore, what is useful to one speaker is highly likely, if it is not tightly context-bound, to be useful to other speakers as well. I believe that these three evolutionary concepts may inform linguistic theories across the board.

As my final point, I would like to refer back to the questions of creativity raised in the opening chapter. First, I believe that we can lay to rest the question of whether the creators of Tok Pisin or other contact languages are more creative than speakers of "ordinary" languages (whatever those may be). They employ the same processes speakers of other languages do, they just operate under different circumstances. This is not to say that speakers of Tok Pisin, past and present, are not creative per se: they are, just as we are. They overcome communicative challenges that present themselves to them with as much success that speakers of any other language do, with a mixture of relying on and recombining existing constructions and introducing new constructions as is necessary.

As we have seen, most of the constructions analysed above do not depend on high-agency processes, with speakers consciously creating their form and function. Does this mean, however, that they are not creative? This is where we get back to the notion of linguistic creativity as an integral part and subprocess of linguistic innovation as defined in the beginning. This indicates that it is not necessary for all innovation to be creative in the sense that it is actively created. Language is inherently a tool for communication: as long as it fulfils its purpose, in familiar circumstances as well as in new ones, it is not necessary for the solutions speakers find to be particularly creative: being innovative is sufficient.

All languages, then, possess the inherent capability for their speakers to be creative in recombining their structure and introducing new material, consciously or unconsciously. Which path they choose, or which path the invisible hand seems to follow, is irrelevant. As every speaker can be creative, every speaker can be innovative. In the word of Ronald Carter, "[l]inguistic creativity is not simply a property of exceptional people but an exceptional property of all people" (2004, 13). Since all people - or speakers - share the property of creativity, it should not come as a surprise that their creations, i.e. languages, reflect this creativity.

Glossing Rules

Gloss	Meaning	Gloss	Meaning
1PL-DUAL	First person dual	FUT	Future
1PL-EXCL	First person plural exclusive	GEN	Genitive
1PL-EXCL-MULT	First person multiple-exclusive	HAB	Habitual
1PL-INC	First person plural inclusive	INCH	Inchoative
1SG	First person singular	ING	Ingressive
2SG	Second person singular	INT	Intensifier
2PL	Second person plural	NEG	Negator
3SG	Third person singular	NEG-PERM	Negated permissive
3PL	Third person plural	NOM	Nominative
ABL	Abilitative	NUM	Numeral
ADV	Adverbial	PL	Plural
ADV-L	Locative adverbial	PREP	Preposition
ADV-T	Time adverbial	PREP-C	Comitative preposition
AGR	Agreement	PREP-D	Directional preposition
ANT	Anterior	PREP-F	Final preposition
ART	Article	PREP-I	Instrumental preposition
ATT	Attemptative	PREP-L	Locative preposition
COMP	Complementizer	PREP-P	Possessive preposition
COMPL	Completive	PREP-T	Temporal preposition
CONC	Concessive	PRN	Pronoun
COND	Conditional	PRN-INT	Intensifying Pronoun
CONJ	Conjunction	PROG	Progressive
CONJ-CONC	Concessive Conjunction	PROH	Prohibitive
DEF	Definite	PST	Past
DEM	Demonstrative	REFL	Reflexive
DEON	Deontic	REL	Relativizer
DT	Determiner	SEQ	Sequential
EMPH	Emphatic	TOP	Topic
EXCL	Exclusive	TR	Transitive
FOC	Focalizer	VB	Verb

Table 8.1: Glossing rules

Bibliography

Ahearn, Laura M. (2001). Language and Agency. *Annual Review of Anthropology*, 30:109–137.

Aitchison, Jean (1989). Spaghetti junctions and recurrent routes. *Lingua*, 77(2):151–171.

Alessandro Duranti (2007). Agency in Language: 20. In *A Companion to Linguistic Anthropology*, pages 449–473. Wiley-Blackwell.

Andersen, Roger W. (1983). *Pidginization and creolization as language acquisition: Roger W. Andersen, editor.* Newbury House, Rowley, Mass.

Arends, Jacques and Bruyn, Adrienne (1994). Gradualist and developmental hypotheses. In Arends, Jacques, Muysken, Pieter, and Smith, Norval, editors, *Pidgins and Creoles: An introduction*, volume 15 of *Creole language library*, pages 111–120. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Australian Government Department of the Environment, Water, Heritage and the Arts (2011). Sapos iu no save gud abaotim diswan—faedem aot!

Bailey, Laura Rudall (2013). *The syntax of question particles.* Doctoral, University of Newcastle Upon Tyne, Newcastle Upon Tyne.

Baker, Philip (1993). Australian influence on Melanesian Pidgin English. *Te Reo*, 36:3–67.

Baker, Philip (1994). Creativity in creole genesis. In Adone, Dany and Plag, Ingo, editors, *Creolization and language change*, Linguistische Arbeiten. Max Niemeyer Verlag, Tübingen.

Baker, Philip and Huber, Magnus (2001). Atlantic, Pacific, and world-wide features in English-lexicon contact languages. *English World-Wide*, 22(2):157–208.

Bartens, Angela (2013a). Creoles. In Bakker, Peter and Matras, Yaron, editors, *Contact languages: A comprehensive guide*, Language contact and bilingualism. De Gruyter Mouton, Boston and Berlin.

Bartens, Angela (2013b). Nicaraguan Creole English. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 115–126. Oxford University Press, Oxford.

Bartens, Angela (2013c). San Andres Creole English. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 101–114. Oxford University Press, Oxford.

Beckner, Clay, Blythe, Richard, Bybee, Joan, Christiansen, Morten H., Croft, William, Ellis, Nick C., Holland, John, Ke, Jinyun, Larsen-Freeman, Diane, and Schoenemann, Tom (2009). Language is a Complex Adaptive System: Position Paper. *Language Learning*, 59(Suppl. 1):1–26.

Bellama, David (1983). *An Introduction to Cameroonian Pidgin: Student's Book.* Peace Corps, Washington, D.C., 2nd edition.

Bennett, Judith A. (1987). *Wealth of the Solomons: A history of a Pacific archipelago, 1800-1978*, volume 3 of *Pacific islands monograph series.* University of Hawaii Press, Honolulu (T.H.).

Bickerton, Derek (1977). Pidginization and creolization: Language acquisition and language universals. *Pidgin and creole linguistics*, pages 46–49.

Bickerton, Derek (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences*, 7(02):173.

Bickerton, Derek (1990). *Language and Species.* University of Chicago Press, Chicago.

Blake, Barry J. (2001). *Case.* Cambridge textbooks in linguistics. Cambridge University Press, Cambridge and New York, 2nd ed. edition.

Bloomfield, Leonard (1933). *Language.* Henry Holt and Co, New York.

Brenninkmeyer, Leo (1924). Einführung ins Pidgin-Englisch: ein Versuch.

Brinton, Laurel J. and Traugott, Elizabeth Closs (2005). *Lexicalization and Language Change: Research Surveys in Linguistics.* Cambridge University Press, Cambridge.

Brown, Keith (2008). *Concise Encyclopedia of Languages of the World.* Concise Encyclopedias of Language and Linguistics. Elsevier professional, s.l., 1. aufl. edition.

Bruyn, Adrienne (2009). Grammaticalization in pidgins and creoles. In Kouwenberg, Silvia and Singler, John Victor, editors, *The Handbook of Pidgin and Creole Studies*, pages 385–410. Wiley-Blackwell, Malden, Mass.

Burkhardt, Richard W. (2013). Lamarck, evolution, and the inheritance of acquired characters. *Genetics*, 194(4):793–805.

Bybee, Joan L., Perkins, Revere D., and Pagliuca, William (1996). *The evolution of grammar: Tense, aspect, and modality in the languages of the world.* Univ. of Chicago Press, Chicago, [nachdr., 2. pr.] edition.

Bybee, Joan L. and Slobin, Dan I. (1982). Rules and Schemas in the Development and Use of the English past Tense. *Language*, 58(2):265.

Carter, Ronald (2004). *Language and creativity: The art of common talk.* Routledge, London and New York.

Cassidy, Frederic G. (1971). Tracing the pidgin element in Jamaican Creole (with notes on method and the nature of pidgin vocabularies). In Hymes, Dell, editor, *Pidginization and creolization of languages: Proceedings of a conference held at the University of the West Indies, Mona, Jamaica, April 1968.* Cambridge Univ. Pr, Cambridge.

Churchill, William (1911). *Beach-La-Mar: The Jargon Or Trade Speech Of the Western Pacific.* Carnegie Institution of Washington.

Cillizza, Chris and Blake, Aaron (March 26th, 2012). President Obama embraces 'Obamacare' label. But why? *Washington Post.*

Clark, Donald H. (1955). Pidgin English: South Pacific polyglot. *Pacific Discovery*, 8(5):8–12.

Clark, Ross (1979). In Search of Beach-la-Mar: Towards a History of Pacific Pidgin English. *Te Reo*, 22:3–64.

Crabtree, William Arthur (1902). *Elements of Luganda Grammar: Together with Exercises and Vocabulary.* Society for Promoting Christian Knowledge, London.

Croft, William (2000). *Explaining language change: An evolutionary approach.* Longman linguistics library. Longman, Harlow.

Croft, William (2006). *Typology and universals.* Cambridge textbooks in linguistics. Cambridge Univ. Press, Cambridge, 2. ed., 3. printing edition.

Crowley, Terry (1990). *Beach-la-Mar to Bislama: The emergence of a national language in Vanuatu.* Oxford Studies in Language Contact. Clarendon Press, Oxford and New York.

Crowley, Terry (1998). The Bislama lexicon before the first world war: Written attestations. In Mühlhäusler, Peter, editor, *Papers in Pidgin and Creole Linguistics*, Pacific linguistics Series A, Occasional papers. Dept. of Linguistics Research School of Pacific Studies Australian Nat. Univ, Canberra.

Crowley, Terry (2004). *Bislama reference grammar*, volume 31 of *Oceanic linguistics special publication.* Univ. of Hawai' Press, Honolulu.

Crowley, Terry (2005). Bislama (Vanuatu).

Crowley, Terry, Lynch, John, and Ross, Malcolm (2013). *The Oceanic Languages.* Routledge Language Family Series. Taylor and Francis, Hoboken.

Crystal, David (1987). *The Cambridge encyclopedia of language.* Cambridge Univ. Press, Cambridge, reprinted. edition.

Cutler, A. and Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and speech*, 20(1):1–10.

de Reuse, Willem J. (1982). Remarks on the Lakhota Enclitics. *Kansas Working Papers in Linguistics.*

de Smet, Hendrik (2009). Analysing reanalysis. *Lingua*, 119(11):1728–1755.

de Young, Karen (October 7th, 2014). Is it a 'war'? an 'armed conflict'? why words matter in the u.s. fight vs. the islamic state. *Washington Post.*

DeCamp, David (1977). The development of pidgin and creole studies. *Pidgin and creole linguistics*, pages 3–20.

DeGraff, Michel (2001). Creolization, language change, and language acquisition: an epilogue. In DeGraff, Michel, editor, *Language creation and language change: Creolization, diachrony, and development*, Learning, development, and conceptual change, pages 473–543. MIT Press, Cambridge, Mass.

Delancey, Scott (1997). Mirativity: The grammatical marking of unexpected information. *Linguistic Typology*, 1(1).

Detges, Ulrich and Waltereit, Richard (2002). Grammaticalization vs. re-analysis: A semantic-pragmatic account of functional change in grammar. *Zeitschrift für Sprachwissenschaft*, 21(2):151–195.

Devonish, Hubert S. and Thompson, Dahlia (2010). *A concise grammar of Guyanese Creole (Creolese)*, volume 80 of *Languages of the world Materials*. LINCOM Europa, München.

Dutton, Thomas Edward (1981). *Conversational New Guinea pidgin*, volume No.12 of *Pacific linguistics Series D, Special publications*. Australian National Univ. Dep. of Linguistics Research School of Pacific Studies, Canberra, repr. der ausg. von 1973 edition.

Dutton, Thomas Edward and Thomas, Dicks (1999). *A new course in Tok Pisin (New Guinea Pidgin)*, volume no. 2 of *Publication / Languages for intercultural communication in the Pacific area project of the Australian Academy of the Humanities*. Dep. of Linguistics Research School of Pacific and Asian Studies Australian National Univ, Canberra, Australia, reprinted. edition.

Eberl, Martin (2019). Supply, demand and...what? Why some features don't get borrowed. In Bader, Katharina, Barberio, Teresa, Huber, Elisabeth, Lasch, Sebastian, Matoshi, Veton, Moser, Ann-Marie, and Robins, Jenny, editors, *JournaLIPP 6*.

Emerine, Rachel Dorlene (2009). *Schools, marriage, and the endangerment of the Nagu: Language in the Solomon Islands*. Ma thesis, California State University, Long Beach, CA.

Engelberg, Stefan and Stolberg, Doris (2017). The Influence of German on the Lexicon of Tok Pisin. *Journal of the Linguistic Society of Papua New Guinea*, Special Issue(Special Issue):27–64.

Escure, Geneviève (2013). Belizean Creole. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 92–100. Oxford University Press, Oxford.

Faraclas, Nicolas (2013). Nigerian pidgin. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 176–184. Oxford University Press, Oxford.

Farquharson, Joseph T. (2013). Jamaican. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 81–91. Oxford University Press, Oxford.

Field, Fredric W. and Comrie, Bernard (2002). *Linguistic borrowing in bilingual contexts*, volume 62 of *Studies in language companion series*. Benjamins, Amsterdam.

Finney, Malcolm Awadajin (2013). Krio. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 157–166. Oxford University Press, Oxford.

Franken, Robert E. (1998). *Human Motivation.* Brooks/Cole Publishing Company, London, 4 edition.

Franklin, Karl James, Kerr, Harland B., and Beaumont, Clive H. (1974). *Tolai language course*, volume 7 of *Language data / Asian-Pacific series*. Summer Institute of Linguistics, Huntington Beach, Calif., 3. ed. edition.

Gast, Volker and König, Ekkehard (2002). Reflexive pronouns and other uses of 'self'-forms in English. *Zeitschrift fur Anglistik und Amerikanistik*, 50.

Gast, Volker and van der Auwera, Johan (2012). What is 'contact-induced grammaticalization'? Examples from Mayan and Mixe-Zoquean languages. In Wiemer, Björn, Wälchli, Bernhard, and Hansen, Björn, editors, *Grammatical replication and borrowability in language contact*, Trends in linguistics. Studies and monographs, 1861-4302. De Gruyter Mouton, Berlin.

Geoff P. Smith and Jeff Siegel (2013). Tok Pisin structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber, editors, *Atlas of Pidgin and Creole Language Structures Online.* Max Planck Institute for Evolutionary Anthropology, Leipzig.

Glare, P.G.W., University of Oxford, and Oxford University Press (1982). *Oxford Latin Dictionary.* Oxford Latin Dictionary. Clarendon Press.

Goldin-Meadow, Susan (2004). Getting a handle on language creation. In Haspelmath, Martin, editor, *Coordinating Constructions*, volume 53 of *Typological studies in language*, pages 343–374. John Benjamins Publishing Company, Amsterdam.

Goury, Laurence and Migge, Bettina (2003). *Grammaire du nengee: Introduction aux langues aluku, ndyuka et pamaka.* Didactiques. IRD, Paris.

Green, Lisa (2013). African American English. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 148–156. Oxford University Press, Oxford.

Greenhill, Simon J., Hua, Xia, Welsh, Caela F., Schneemann, Hilde, and Bromham, Lindell (2018). Population Size and the Rate of Language Evolution: A Test Across Indo-European, Austronesian, and Bantu Languages. *Frontiers in Psychology*, 9:317–334.

Hackert, Stephanie (2004). *Urban Bahamian Creole: System and variation*, volume v. G32 of *Varieties of English around the world*. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Hackert, Stephanie (2013). Bahamian Creole. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 127–138. Oxford University Press, Oxford.

Hall, R. A. (1966). *Pidgin and creole languages.* Cornell University Press.

Hall, JR., Robert A. (1943a). *Melanesian Pidgin English: Grammar, Texts, Vocabulary.* Linguistic Society of America, Madison, Wisconsin.

Hall, JR., Robert A. (1943b). The vocabulary of melanesian pidgin english. *American Speech*, 18(3):192–199.

Haspelmath, Martin (1998). Does grammaticalization need reanalysis? *Studies in Language*, 22(2):315–351.

Haspelmath, Martin (2004). 1. coordinating constructions. In Haspelmath, Martin, editor, *Coordinating Constructions*, volume 58 of *Typological studies in language*, pages 3–39. John Benjamins Publishing Company, Amsterdam.

Haspelmath, Martin (2013). Nominal and verbal conjunction. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages. Oxford University Press, Oxford.

Heath, Jeffrey (1978). *Linguistic diffusion in Arnhem Land.* Australian Institute of Aboriginal Studies, Canberra.

Heine, Bernd and Kuteva, Tania (2002). *World Lexicon of grammaticalization.* Cambridge University Press, Cambridge.

Heine, Bernd and Kuteva, Tania (2003). On contact-induced grammaticalization. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 27(3):529–572.

Heine, Bernd and Traugott, Elizabeth Closs, editors (1991). *Approaches to grammaticalization: Volume 1: Focus on theoretical and methodological issues*, volume v. 19, 1 of *Typological studies in language.* John Benjamins Pub. Co, Amsterdam.

Hempenstall, Peter J. (1975). The Reception of European Missions in the German Pacific Empire: The New Guinea Experience. *The Journal of Pacific History*, 10(1):46–64.

Hickey, Raymond (2014). *A dictionary of varieties of English.* Wiley-Blackwell, Chichester, West Sussex, U.K, first edition.

Holm, John A. (1988). *Pidgins and creoles. Volume 1: Theory and structure.* Cambridge University Press, Cambridge.

Holm, John A. (1994). *Pidgins and creoles.* Cambridge language surveys. Cambridge Univ. Press, Cambridge, reprinted edition.

Höltker, Georg (1945). Das Pidgin-Englisch als sprachliches Missionsmittel in Neuguinea. *Neue Zeitschrift für Missionswissenschaft*, 1:44–63.

Hopper, Paul J. and Traugott, Elizabeth Closs (1993). *Grammaticalization.* Cambridge textbooks in linguistics. Cambridge Univ. Press, Cambridge.

Huber, Magnus (2013). Ghanaian Pidgin English. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 167–175. Oxford University Press, Oxford.

Hudson, W. J. (1965). Australia's experience as a mandatory power. *Australian Outlook*, 19(1):35–46.

Huebner, Thom and Horoi, Stephen Rex (1979). *Solomon Islands Pijin: Grammar Handbook.* Peace Corps Language Handbook Series. ACTION/Peace Corps, Brartleboro, Vermont.

Inoue, Aya (2007). From Exclusive Particle to Adversative Conjunction: A Study on the Particle tasol in Tok Pisin. *Annual Meeting of the Berkeley Linguistics Society*, 33(1):233.

Jacobs, Thomas Jefferson (1844). *Scenes, Incidents and Adventures in the Pacific Ocean.* Harpers and Brothers, New York.

Jakobson, Roman (1936). Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutungen der russischen Kasus. In *Traveux Du Cercle Linguistique de Prague 6*, pages 240–288. n.p, Prague.

Jespersen, Otto (1922). *Language: its nature, development and origin.* H. Holt, London and G. Allen & Unwin ltd.

Johnson, Robert Keith (1977). Administration and Language Policy in Papua New Guinea. In Wurm, Stephen A., editor, *New Guinea area languages and language study*, Pacific linguistics / C. Dep. of Linguistics, Research School of Pacific Studies, Australian National Univ, Canberra.

Jones, Daniel and Roach, Peter, editors (2009). *Cambridge English pronouncing dictionary: Spoken pronunciation for every word - now american english too.* Cambridge Univ. Press, Cambridge, new 17th ed., 4th print edition.

Jourdan, Christine (2004). Solomon Islands Pijin: morphology and syntax. In Kortmann, Bernd, Schneider, Edgar W., Burridge, Kate, Mesthrie, Rajend, and Upton, Clive, editors, *A Handbook of Varieties of English.* Mouton de Gruyter, Berlin and New York.

Jourdan, Christine (2008). Language repertoires and the middle class in urban Solomon Islands. In Meyerhoff, Miriam and Nagy, Naomi, editors, *Social lives in language - sociolinguistics and multilingual speech communities ; celebrating the work of Gillian Sankoff*, Impact, pages 43–67. Benjamins, Amsterdam.

Jourdan, Christine and Keesing, Roger M. (1997). From Fisin to Pijin: Creolization in Process in the Solomon Islands. *Language in Society*, 26(3):401–420.

Kaufman, James C. and Beghetto, Ronald A. (2009). Beyond Big and Little: The Four C Model of Creativity. *Review of General Psychology*, 13(1):1–12.

Kaufman, James C. and Sternberg, Robert J. (2010). *The Cambridge Handbook of Creativity.* Cambridge University Press, Cambridge.

Keesing, Roger M. (1988). *Melanesian Pidgin and the Oceanic substrate.* Stanford University Press, Stanford.

Kegl, Judy and McWhorter, John H. (1997). Perspectives on an emerging language. In Clark, Eve, editor, *Proceedings of the Twenty-Eighth Annual Child Language Research Forum*, pages 15–38.

Keller, Rudi (1994). *On language change: The invisible hand in language.* Routledge, London and New York.

Kempkes, Lambert (1968). *Catholic Mission Involvement with Education in Papua-New Guinea.* Master of arts, Loyola University, Chicago.

Kersting, Karen (2003). What exactly is creativity? Psychologists continue their quest to better understand creativity. *Monitor on Psychology*, 34(10):40.

Klein, Thomas B. (2013). Gullah. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 139–147. Oxford University Press, Oxford.

Landtman, Gunnar (1918). The Pidgin-English of British New Guinea. *Neuphilologische Mitteilungen*, 19(7/8):62–74.

Langacker, Ronald W. (1977). Syntactic reanalysis. *Mechanisms of syntactic change*, 58.

Lawrence, Peter (1964). *Road belong Cargo: A study of the Cargo movement in the Southern Madang District New Guinea.* Manchester University Press, University of Michigan.

Lehmann, Christian (2004). Theory and method in grammaticalization [Theorie und Methode in der Grammatikalisierung]. *Zeitschrift für germanistische Linguistik*, 32(2):231.

Lehmann, Christian (2015). *Thoughts on grammaticalization*, volume 1 of *Classics in Linguistics.* Language Science Press, Berlin, Germany, 3rd edition edition.

Lever, Robert John Aylwin Wallace (1964). Whales and whaling in the western pacific. *South Pacific Bulletin*, April.

Lim, Lisa and Ansaldo, Umberto (2013). Singlish structure dataset. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages. Oxford University Press, Oxford.

Lupyan, Gary and Dale, Rick (2010). Language structure is partly determined by social structure. *PloS one*, 5(1):e8559.

Luraghi, Silvia and Narrog, Heiko (2014). *Perspectives on Semantic Roles*, volume 106 of *Typological studies in language*. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Maggio, Rodolfo (2018). "According to Kastom and According to Law": "Good Life" and "Good Death" in Gilbert Camp, Solomon Islands. In Gregory, Chris and Altman, Jon, editors, *The Quest for the Good Life in Precarious Times: Ethnographic Perspectives on the Domestic Moral Economy*, pages 57–86. ANU Press.

Maria Polinsky and Eric Potsdam (in press). Austronesian syntax. In *Oceania*. Mouton, Berlin.

Markey, Thomas L. (1982). Afrikaans: Creole or non-creole. *Zeitschrift für Dialektologie und Linguistik*, 49(2).

Matras, Yaron (1998). Utterance modifiers and universals of grammatical borrowing. *Linguistics*, 36(2).

Matras, Yaron (2007). The borrowability of structural categories. In Sakel, Jeanette and Matras, Yaron, editors, *Grammatical borrowing in cross-linguistic perspective*, Empirical approaches to language typology, pages 31–74. Mouton de Gruyter, New York.

Matras, Yaron (2009). *Language contact*. Cambridge textbooks in linguistics. Cambridge University Press, Cambridge.

Matthews, Stephen and Li, Michelle (2013). Chinese Pidgin English. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 206–213. Oxford University Press, Oxford.

Mayr, Ernst and Diamond, Jared M. (2001). *The birds of northern Melanesia: Speciation, ecology, & biogeography.* Oxford University Press, New York.

McMahon, April M. S. (2002). *Understanding language change.* Cambridge Univ. Press, Cambridge.

McWhorter, John H. (1995). The scarcity of Spanish-based creoles explained. *Language in Society*, 24(2):213–244.

McWhorter, John H. (2018). *The creole debate.* Cambridge University Press, Cambridge.

McWhorter, John H. and Good, Jeff (2012). *A grammar of Saramaccan Creole*, volume 56 of *Mouton grammar library.* De Gruyter Mouton, Berlin.

Mead, Margaret (1931). Talk-boy. *Asia*, 31:141–151.

Mead, Margaret (1956). *New lives for old.* William Morrow, New York.

Meakins, Felicity (2011). *Case-marking in contact: The development and function of case morphology in Gurindji Kriol*, volume v. 39 of *Creole language library.* John Benjamins Pub. Co, Amsterdam and Philadelphia.

Meyerhoff, Miriam (2013). Bislama. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 223–231. Oxford University Press, Oxford.

Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors (2013). *The survey of pidgin and creole languages.* The atlas and survey of pidgin and creole languages. Oxford University Press, Oxford.

Michaelis, Susanne and Rosalie, Marcel (2000). Polysémie et cartes sématiques: Le relateur (av)ek en créole seychellois.

Mihalic, Francis (1957). *Grammar and dictionary of Neo-Melanesian.* Mission Press, Techny, Illinois.

Mihalic, Francis (1977). Policies and Experiences in the Catholic Mission in the Field of Teaching in Pidgin. In Wurm, Stephen A., editor, *New Guinea area languages and language study*, Pacific linguistics / C, pages 653–659. Dep. of Linguistics, Research School of Pacific Studies, Australian National Univ, Canberra.

Mohammed, Mohammed Abdulla (2001). *Modern Swahili grammar.* East African Education Publ, Nairobi.

Monash University Library (2009). Tok bilog gavman.

Mosel, Ulrike (1980). *Tolai and Tok Pisin: The influence of the substratum on the development of New Guinea Pidgin*, volume 73 of *Pacific Linguistics Series B, Monographs.* Dept. of Linguistics Research School of Pacific Studies Australian National Univ, Canberra.

Mosel, Ulrike (1984). *Tolai syntax and its historical development.* Department of Linguistics, Research School of Pacific Studies, Australian National University, Canberra.

Mufwene, Salikoko S. (1986). The Universalist and Substrate Hypotheses Complement One Another. In Muysken, Pieter and Smith, Norval, editors, *Substrata versus Universals in Creole Genesis*, volume 1 of *Creole language library*, page 129. John Benjamins Publishing Company, Amsterdam.

Mufwene, Salikoko S. (2008). *Language evolution: Contact, competition and change.* Continuum, London and New York.

Mufwene, Salikoko S., Pellegrino, François, and Coupé, Christophe (2017). *Complexity in language: Developmental and evolutionary perspectives.* Cambridge Approaches to Language Contact. Cambridge University Press, Cambridge.

Mugler, France, editor (1996). *Pacific Languages in Education.* Institute of Pacific Studies, Suva.

Mühleisen, Susanne (2013). Trinidad English Creole. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages. Oxford University Press, Oxford.

Mühlhäusler, P. (1974). *Pidginization and simplification of language.* Pacific linguistics. Series B, Monographs. Dept. of Linguistics, Research School of Pacific Studies, Australian National University.

Mühlhäusler, P. (1979). *Growth and structure of the lexicon of New Guinea Pidgin.* Pacific linguistics. Dept. of Linguistics, Research School of Pacific Studies, Australian National University.

Mühlhäusler, Peter (1982). Tok Pisin in Papua New Guinea. In Bailey, Richard W. and Görlach, Manfred, editors, *English as a world language*, pages 439–467. University of Michigan Press, Ann Arbor.

Mühlhäusler, Peter (1990). On the origins of the predicate marker in Tok Pisin. In Verhaar, John W. M., editor, *Melanesian Pidgin and Tok Pisin: Proceedings of the First International Conference of Pidgins and Creoles in Melanesia*, Studies in language companion series, pages 235–249. John Benjamins Pub. Co, Amsterdam and Philadelphia.

Mühlhäusler, Peter, editor (1997). *Pidgin and Creole Linguistics.* University of Westminster Press, London.

Mühlhäusler, Peter (2003). *Tok Pisin texts from the beginning to the present*, volume 9 of *Varieties of English around the world : Text series.* Benjamins, Amsterdam.

Mühlhäusler, Peter (2013). Norf'k. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 232–240. Oxford University Press, Oxford.

Muysken, Pieter (1981). Creole tense/mood/aspect systems: the unmarked case? In Muysken, Pieter, editor, *Generative studies on Creole languages*, Studies in Generative Grammar, pages 181–199. De Gruyter Mouton, Berlin and Boston.

Muysken, Pieter (2008). *Functional categories*, volume 117 of *Cambridge studies in linguistics.* Cambridge Univ. Press, Cambridge, 1. publ edition.

Nettle, Daniel (2012). Social scale and structural complexity in human languages. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1597):1829–1836.

Newmeyer, Frederick J. (1998). *Language form and language function.* MIT Press, Cambridge, Mass.

Orken, M. (1954). Pidgin English and the Tolais of New Britain. *South Pacific Bulletin*, 7:863.

Payne, Thomas Edward (1997). *Describing morphosyntax: A guide for field linguists.* Cambridge Univ. Press, Cambridge, 1. publ edition.

Prescod, Paula (2013). Vincentian Creole. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 70–80. Oxford University Press, Oxford.

Ramat, Anna Giacalone and Mauri, Caterina (2011). The grammaticalization of coordinating interclausal connectives. In Narrog, Heiko and Heine, Bernd, editors, *The Oxford handbook of grammaticalization*, Oxford handbooks in linguistics, pages 656–667. Oxford Univ. Press, Oxford.

Raviv, Limor, Meyer, Antje, and Lev-Ari, Shiri (2018). The role of community size in the emergence of linguistic structure. In *Proceedings of the 12th International Conference on the Evolution of Language (Evolang12)*. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

Reali, Florencia, Chater, Nick, and Christiansen, Morten H. (2018). Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society of London B: Biological Sciences*, 285(1871).

Regional Assistance Mission to Solomon Islands (2011). Tok Pijin: A Quick Guide To Speaking Pijin.

Reinecke, John E. (1937). *Marginal languages: A sociological survey of the creole languages and trade jargons*. Ph.d. thesis, Yale University.

Riehl, Claudia Maria (2015). *Mehrsprachigkeit: Eine Einführung*. WBG - Wissenschaftliche Buchgesellschaft, Darmstadt.

Roberge, Paul T. (2009). The creation of pidgins as a possible window on language evolution. In Botha, Rudolf P. and Swart, Henriëtte de, editors, *Language evolution: The view from restricted linguistic systems*, Occasional series / Landelijke Onderzoekschool Taalwetenschap, pages 101–138. LOT, Utrecht.

Roberts, Sarah J. and Bresnan, Joan (2008). Retained inflectional morphology in pidgins: A typological study. *Linguistic Typology*, 12(2):571.

Romaine, Suzanne (1992a). *Language, education, and development: Urban and rural Tok Pisin in Papua New Guinea*. Oxford Studies in Language Contact. Clarendon, Oxford.

Romaine, Suzanne (1992b). Variability in Tok Pisin phonology: "Did you say 'pig' or 'fig'?". In Rissanen, Matti, Ihalainen, Ossi, Nevalainen, Terttu, and Taavitsainen, Irma, editors, *History of Englishes*, volume 8, page 86. De Gruyter Mouton, Berlin, Boston.

Rusling, James Fowler (1874). *Across America: or The Great West and the Pacific Coast*. Sheldon and Company, New York.

Ryan, Peter (1972). *Encyclopaedia of Papua and New Guinea.* Melbourne University Press and University of Papua and New Guinea, Carlton and Port Moresby.

Sack, P. G. (1972). Finsch, Otto (1839-1917).

Sankoff, Gillian, editor (1980). *The Social Life of Language.* University of Pennsylvania Press, Philadelphia.

Sankoff, Gillian (1993). Focus in Tok Pisin. In Byrne, Francis and Winford, Donald, editors, *Focus and grammatical relations in creole languages*, Creole language library 0920-9026, pages 117–140. J. Benjamins, Amsterdam and Philadelphia.

Sankoff, Gillian (1999). The evolution of focus particles and related morphemes in Tok Pisin, Solomon Island Pidgin and Bislama.

Sankoff, Gillian (2010). Using the Future to Explain the Past. In Bickerton, Derek, Byrne, Francis, and Huebner, Thom, editors, *Development and structures of Creole languages: Essays in honor of Derek Bickerton*, Creole language library, pages 61–74. J. Benjamins Pub, Amsterdam and Philadelphia.

Sankoff, Gillian and Brown, Penelope (1976). The origins of syntax in discourse: A case study of tok pisin relatives. *Language*, 52(03):631–666.

Sapir, Edward (1921). *Language: An introduction to the study of speech.* Harcourt, Brace and Co, New York.

Schleicher, August (1863). *Die darwinische Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Dr. Ernst Häckel.* Böhlau, Weimar.

Schleicher, August (1869). *Darwinism tested by the science of language: Translated by Alexander V.W. Bikkers.* J.C. Hotten, London.

Schmid, Hans-Jörg (2015). A blueprint of the entrenchment-and- conventionalization model. *Yearbook of the German Cognitive Linguistics Association*, 3(1).

Schröder, Anne (2013). Cameroon Pidgin English. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 185–193. Oxford University Press, Oxford.

Schuchardt, Hugo (1914). *Die Sprache der Saramakkaneger in Surinam.* Johannes Müller, Amsterdam.

Schultze-Berndt, Eva, Meakins, Felicity, and Angelo, Denise (2013). Kriol. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 241–251. Oxford University Press, Oxford.

Sebba, Mark (1987). *The Syntax of Serial Verbs: An investigation into serialisation in Sranan and other languages*, volume No. 2 of *Creole language library.* John Benjamins Pub. Co.

Senghas, Ann and Coppola, Marie (2001). Children Creating Language: How Nicaraguan Sign Language Acquired a Spatial Grammar. *Psychological Science*, 12(4):323–328.

Shineberg, Dorothy (1999). *The people trade: Pacific Island laborers and New Caledonia, 1865-1930*, volume 16 of *Pacific islands monograph series.* Center for Pacific Islands Studies School of Hawaiian Asian and Pacific Studies University of Hawai'i, Honolulu.

Shineberg, Dorothy (2014). *They came for sandalwood: A study of the sandalwood trade in the south-west Pacific, 1830-1865.* Pacific studies series. University of Queensland Press, St Lucia, Queenland, Australia.

Siegel, Jeff (2000). Substrate influence in Hawaii Creole English. *Language in Society*, 29(2):197–236.

Siegel, Jeff (2008). *The emergence of pidgin and Creole languages.* Oxford linguistics. Oxford University Press, Oxford.

Siegel, Jeff (2009). *Language contact in a plantation environment: A sociolinguistic history of Fiji*, volume 5 of *Studies in the social and cultural foundations of language.* Cambridge Univ. Press, Cambridge, digitally printed version edition.

Siegel, Jeff and Smith, Geoff P. (2013). Tok Pisin. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 214–222. Oxford University Press, Oxford.

Smeall, Christopher (1973). Quantitative Analysis of Variation: i in Tok Pisin. *Annual Meeting of the Berkeley Linguistics Society*, 1.

Smith, Adam (2007). *An Inquiry into the Nature and Causes of the Wealth of Nations.* MetaLibri, New York.

Smith, Geoff P. (2002a). Corpus assembled for Growing up with Tok Pisin. Contact, creolization, and change in Papua New Guinea's national language. London: Battlebridge.

Smith, Geoff P. (2002b). *Growing up with Tok Pisin: Contact, creolization, and change in Papua New Guinea's national language.* Battlebridge Publications, London.

Speiser, Philip (1913). *Südsee, Urwald, Kannibalen.* Vogtländer, Leipzig.

Stanley, Peter (2003). New Guinea Offensive. *Wartime*, 23.

Stassen, Leon (2000). AND-languages and WITH-languages. *Linguistic Typology*, 4(1):1–54.

Steinbauer, Friedrich. (1969). *Concise dictionary of New Guinea pidgin (Neo-Melanesian) with translation in English and German.* Kristen Pres, Madang, New Guinea.

Stevens, David (2001). The naval campaign for New Guinea. *Journal of the Australian War Memorial*, 34.

Taylor, Douglas (1971). Grammatical and Lexical Affinities of Creoles. In Hymes, Dell, editor, *Pidginization and creolization of languages: Proceedings of a conference held at the University of the West Indies, Mona, Jamaica, April 1968.* Cambridge Univ. Pr, Cambridge.

Tintazul (2014). The geographical extent of Melanesia: Map of Melanesia based on the United Nations geoscheme M49 coding classification devised by the United Nations Statistics Division: Svg format.

Tolkien, John Ronald Reuel (1931). A secret vice.

Traugott, Elizabeth Closs (2013). Grammaticalization. In Luraghi, Silvia and Bubeník, Vít, editors, *Bloomsbury companion to historical linguistics*, Bloomsbury companions, pages 269–283. Bloomsbury, London.

Traugott, Elizabeth Closs and König, Ekkehard (1991). The Semantics-Pragmatics of Grammaticalization Revisited. In Heine, Bernd and Traugott, Elizabeth Closs, editors, *Approaches to grammaticalization: Volume 1: Focus on theoretical and methodological issues*, Typological studies in language, pages 189–218. John Benjamins Pub. Co, Amsterdam.

Tryon, Darrell T. and Charpentier, Jean-Michel (2004). *Pacific pidgins and creoles: Origins, growth and development*, volume 132 of *Trends in linguistics. Studies and monographs.* Mouton de Gruyter, Berlin and New York.

Ukman, Jason (December 12th, 2011). Ex-blackwater firm gets a name change, again. *Washington Post.*

Vader, John (1972). *New Guinea: The tide is stemmed.* The Pan/Ballantine illustrated history of World War II. Pan/Ballantine, London.

van den Berg, Margot C. and Smith, Norval S.H. (2013). Early Sranan. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 3–14. Oxford University Press, Oxford.

Velupillai, Viveka (2013). Hawai'i Creole. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 252–261. Oxford University Press, Oxford.

Velupillai, Viveka (2015). *Pidgins, Creoles and mixed languages: An introduction*, volume 48 of *Creole language library.* John Benjamins Publishing Company, Amsterdam and Philadelphia.

Verhaar, John W. M., editor (1990). *Melanesian Pidgin and Tok Pisin: Proceedings of the First International Conference of Pidgins and Creoles in Melanesia*, volume 20 of *Studies in language companion series.* John Benjamins Pub. Co, Amsterdam and Philadelphia.

Verhaar, John W. M. (1995). *Toward a reference grammar of Tok Pisin: An experiment in corpus linguistics*, volume 26 of *Oceanic linguistics special publication.* University of Hawaii Press, Honolulu.

Wardhaugh, Ronald and Fuller, Janet M. (2015). *An introduction to sociolinguistics.* Blackwell textbooks in linguistics. John Wiley & Sons Inc, Hoboken, NJ, seventh edition.

Weinreich, Uriel, Labov, William, and Herzog, Marvin I. (1968). Empirical foundations for a theory of language change. In Lehmann, Winfred P. and Malkiel, Yakov, editors, *Directions for historical linguistics: A symposium.* Univ. of Texas Press, Austin.

Winford, Donald (2000). Intermediate creoles and degrees of change in creole formation: The case of Bajan. In Schneider, Edgar Werner and Neumann-Holzschuh, Ingrid, editors, *Degrees of restructuring in Creole languages*, Creole language library, pages 215–246. J. Benjamins, Amsterdam and Philadelphia.

Winford, Donald (2009). *An introduction to contact linguistics*, volume 33 of *Language in society*. Blackwell, Oxford.

Winford, Donald (2012). Creole languages. In Binnick, Robert I., editor, *The Oxford handbook of tense and aspect*, pages 428–457. Oxford Univ. Press, Oxford.

Woolford, Ellen (1981). The developing complementizer system in Tok Pisin. In Muysken, Pieter, editor, *Generative studies on Creole languages*, volume 6 of *Studies in Generative Grammar*, pages 125–139. De Gruyter Mouton, Berlin and Boston.

Wray, Alison and Grace, George W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117(3):543–578.

Wurm, Stephen A. and Mühlhäusler, Peter (1985). *Handbook of Tok Pisin (New Guinea Pidgin)*, volume publication no. 1 of *Languages for intercultural communication in the pacific area project of the Australian Academy of the Humanities*. Dept. of Linguistics, Research School of Pacific Studies, Australian National University, Canberra, A.C.T., Australia.

Yakpo, Kofi (2009). *A grammar of Pichi*. s.l.

Yakpo, Kofi (2013). Pichi. In Michaelis, Susanne, Maurer, Philippe, Haspelmath, Martin, and Huber, Magnus, editors, *The survey of pidgin and creole languages*, The atlas and survey of pidgin and creole languages, pages 194–205. Oxford University Press, Oxford.

Yule, George (1996). *The study of language*. Cambridge Univ. Press, Cambridge, 2nd edition.

Zawada, Britta Edelgard (2009). *Linguistic creativity and mental representation with reference to intercategorial polysemy: PhD dissertation submitted to the University of South Africa*. Phd dissertation, University of South Africa, Pretoria, South Africa.

Zeschel, Arne (2012). *Incipient Productivity: A Construction-Based Approach to Linguistic Creativity*, volume 49 of *Cognitive linguistics research*. Mouton de Gruyter, s.l., 1st edition.

Zgraggen, John (2011). Texts in Tok Pisin (PNG).